UNIWERSYTET WARSZAWSKI
WYDZIAŁ CHEMII

Molecular modelling of thymidylate synthase and rational design of its inhibitors as novel anticancer drugs

Modelowanie molekularne syntazy tymidylanowej i racjonalne projektowanie jej inhibitorów jako leków przeciwnowotworowych nowej generacji

Agata Jurkiewicz

Praca doktorska
wykonana w Pracowni Chemii Kwantowej
Wydziału Chemii Uniwersytetu Warszawskiego
pod kierownictwem

prof. dr hab. Andrzeja Lesia
i
dr Agnieszki K. Bronowskiej

Note

I am particularly grateful to Professor Andrzej Leś and Dr. Agnieszka Bronowska, who believed in me and supported me in this work in so many ways. I would also like to thank Professor Wojciech Rode for long-time collaboration and his helpful comments on the text.

All this would not be possible without the support and great patience of my loved ones. Thank you all (from the bottom of my heart).

Streszczenie

W poszukiwaniu leków przeciwnowotworowych nowej generacji badano potencjalne inhibitory enzymu syntazy tymidylanowej. Opisano szereg etapów komputerowo wspomaganego projektowania leków. Wybrano dwa miejsca docelowe dla poszukiwanych inhibitorów: kieszeń aktywną enzymu oraz kieszeń allosteryczną między podjednostkami białka. Potencjalnie obiecujące związki wybrano w drodze wysokowydajnej procedury przesiewania (przy zastosowaniu metod dokowania molekularnego) dostępnych baz danych leków i proleków, a następnie modyfikację i dalszą selekcję wyników dokowania. Dla wybranych potencjalnych inhibitorów syntazy tymidylanowej, zarówno kompetycyjnych, jak i allosterycznych, przeprowadzono symulacje metodą dynamiki molekularnej w celu oceny dynamiki układu, parametrów wiązania i ułożenia ligandów, jak również wskazania wiodących związków do dalszych badań *in vitro*. Ponadto opisano bibliotekę peptoidów, stworzoną w celu projektowania nowej generacji związków o pożądanych właściwościach peptydomimetycznych. Wykonano również obliczenia metodami mechaniki kwantowej mające na celu wspomaganie badań i syntezy nowych inhibitorów enzymów, w tym związków zawierających bor.

Summary

In search of novel anticancer drugs, putative inhibitors of the enzyme thymidylate synthase were investigated. The dissertation presents several steps of computationally aided drug design. Two targets are described: active site of the enzyme, for competitive inhibitors, and an allosteric pocket at the dimer interface. The potential hits were selected by computational high-throughput screening (molecular docking calculations) of available drug and prodrug databases. The selected compounds were then modified and scored further to indicate potential leads. Molecular dynamics simulations were performed for selected putative inhibitors of thymidylate synthase, both competitive and allosteric, in order to assess their dynamical behaviour, binding properties and arrangement of the ligands, and to select lead compounds for further tests *in vitro*. Moreover, a library of peptoids is described, created with the aim to design novel compounds with the desired peptide-like properties. Furthermore, quantum mechanics calculations were conducted to aid the synthesis and investigation of novel enzyme inhibitors, including boron containing compounds.

Table of contents

Note ... 3
Streszczenie ... 5
Summary ... 5
Table of contents ... 7
Abbreviations ... 8
1 Introduction ... 9
1.1 Drug design ... 9
1.2 Thymidylate synthase ... 10
1.2.1 Structure, dynamics, and conformational changes of TS ... 10
1.2.2 Thymidylate synthase reaction ... 12
1.2.3 Dynamical behaviour of thymidylate synthase ... 12
1.2.4 Thymidylate synthase ligands ... 13
1.2.5 Existing drugs targeting TS ... 15
2 The goals of this work ... 18
3 Methods ... 19
3.1 Molecular docking ... 20
3.2 Molecular dynamics ... 22
3.2.1 Basic concepts of molecular dynamics ... 22
3.2.2 Preparing the MD simulations ... 24
3.2.3 Analysis of the MD simulation ... 25
4 Summary of the results ... 26
4.1 Published paper ... 26
4.2 Non-published work ... 28
4.2.1 Hydroxyl analogues of dUMP ... 28
4.2.2 Virtual screening and docking ... 33
4.2.3 Drug repositioning ... 39
4.2.4 Peptidomimetics ... 40
5 Other work ... 46
5.1 DFT studies of interactions between nucleic acid bases ... 46
5.2 Boron analogues of nucleosides ... 46
5.3 Modification of stavudine ... 47
5.4 Non-enzymatic modifications of thymidylate synthase ... 49
6 Discussion, conclusions and future outlook ... 51
7 Acknowledgements ... 56
8 References ... 56
9 Appendices ... 61
1 Fusaro *et al*., 2010 ... 61
2 Poltev *et al*., 2008 ... 61
3 Poltev *et al*., 2009 ... 61
4 Ruman *et al*., 2009a ... 61
5 Ruman *et al*., 2007 ... 61
6 Ruman *et al*., 2010a ... 61
7 Ruman *et al*., 2009b ... 61
8 Ruman *et al*., 2010b ... 61
9 Dąbrowska-Maś *et al*., 2012 ... 61

Abbreviations

ADMET	adsorption, distribution, metabolism, excretion, toxicity
AM1-BCC	semi-empirical AM1 Hamiltonian Bond Charge Corrections
DFT	Density Functional Theory
DHF	dihydrofolate
DMP	deoxyuridine monophosphate
dUMP	deoxyuridine monophosphate
dTTP	deoxythymidine triphosphate
FDA	Food and Drug Administration (USA)
FdUMP	5-fluorodeoxyuridine monophosphate
FMP	5-fluorodeoxyuridine monophosphate
5-FU	5-fluorouracil
hTS	human thymidylate synthase
MD	molecular dynamics
MM	molecular mechanics
MOD	dUMP-OCH_3
MS	mass spectrometry
MTF	methylenetetrahydrofolate
mTHF	methylenetetrahydrofolate
OCD	dUMP-CH_2OH
OHD	dUMP-OH
PM7*	semi-empirical PM7 Hamiltonian
RESP	restrained electrostatic potential
RMSD	root-mean-square deviation
QM	quantum mechanics

1 Introduction

1.1 Drug design

It takes about 7-12 years to design, optimise, test and introduce a new drug into the market, at a cost of around USD 0.5-2 billion. According to estimations, only one of 8000 compounds tested in animals reaches human testing, and of these, only one out of five is approved for use. Additionally, some drugs already present on the market may be later withdrawn due to serious, sometimes fatal side effects. Alternative approaches that would lower the costs, shorten the time and reduce the need for animal testing while minimising the risks associated with new substances, are needed. Molecular modelling methods can fulfil some of the requirements of the drug design process, and their significance increases with the advancements in computing power, development of new computational methods applied to biological problems, and an increase of the amount of data available in public databases.

The design and introduction of a new drug involves the following steps:

- target identification and validation – establishing the target or targets (such as specified enzymes) for treatment of a given disease
- hit discovery – a "hit" is a result of high throughput screening (e.g. computational docking of a set of compounds from a database to the protein), a molecule with some desired properties towards the established target
- hit-to-lead conversion – selection and preliminary optimisation of the promising ligands in order to find the "lead" compounds, substances that are most likely to be useful in further development
- lead optimisation – modification of the selected lead compound in order to improve its ADMET (adsorption, distribution, metabolism, excretion, and toxicity) properties
- pre-clinical tests – performed on animal models in order to establish pharmacodynamical, pharmacokinetical, and ADMET properties of the drug
- clinical trials – performed on patients, comprise of four phases with the aim to determine the actual therapeutic effects and toxicity of the drug, appropriate therapeutic dosage, side effects, and interactions with other drugs. Clinical tests include also post-marketing surveillance of the long-term effects.

An essential consideration in drug discovery and development are the ADMET properties. A compound may be a strong, selective inhibitor of an enzyme *in silico* and *in vitro*, but it is of limited therapeutical use if it does not cross biological barriers, is not soluble, is unstable or accumulates in the body. For example, peptides can bind well to a receptor, but are prone to proteolytic digestion and therefore difficult to utilise as drugs. There are also traps on the metabolic pathway of the drug – various reactions can lead to

the formation of toxic derivatives or interactions can occur with other drugs or receptors that are not the intended target [Merino *et al.*, 2010].

Several thousand of target proteins are already known, however, effective drugs exist for only a small fraction (about 250) of the diseases for which the molecular mechanism is established. Thymidylate synthase is an example of an enzyme the action of which is well described and competitive (orthosteric) inhibitors exist, but their therapeutic potential is somewhat limited. Within this already known protein, a new target was recently identified [Cardinale *et al*., 2011]: the dimer interface that may adapt new, allosteric inhibitors of the enzyme.

1.2 Thymidylate synthase

The reaction catalysed by human thymidylate synthase (hTS, EC 2.1.1.45) constitutes an essential step of the only pathway for producing *de novo* 2'-deoxythymidine-5'-triphosphate (dTTP) for DNA synthesis. Therefore the enzyme is a good target for anticancer, antifungal and antiviral drugs, inhibiting the growth of rapidly proliferating cells by interfering with TS function. Overexpression of thymidylate synthase is common in many cancers and the tumours may be less sensitive to TS inhibitors than healthy cells. Moreover, the use of inhibitors such as fluoropyrimidines can lead to TS gene amplification and increased levels of the protein. In any case, high level of TS in cancer cells is correlated with poor clinical response to fluoropyrimidines and the rapid progress of the cancer [Li *et al.*, 2001].

The subject of this work is methylenetetrahydrofolate-dependent thymidylate synthase, found in most organisms, including mammals. It is so well conserved through the evolution that bacterial (e.g. *E. coli* and *L. casei*) enzymes may be used as models of mammalian ones. All these TSes share the same overall fold defined by 8 α-helices and 10 β-sheets, with a 6-stranded twisted β-sheets forming the dimer interface, and a deep active site cavity (Figure 1). There is high sequence similarity between bacterial and mammalian TSes in the regions surrounding the active site and dimer interface [Carreras and Santi, 1995].

Some microorganisms produce thymidine via FDTS – flavin-dependent thymidylate synthase-catalysed reaction [Koehn and Kohen, 2009], which is not covered here.

1.2.1 Structure, dynamics, and conformational changes of TS

Thymidylate synthase is a homodimeric protein, molecular weight about 63 kDa, in which both subunits contribute to each of the two active sites. However, only one active site appears to be functional at a time [Świniarska *et al*., 2010]. In the absence of ligands, the active site of TS shows an open conformation allowing the ligands to enter it. The formation of the TS-ligand-cofactor complex (the ternary complex) leads to a conformational change in the protein and almost complete burial of ligands. As indicated by the structures of the complex of TS with dUMP (the binary complex), and by NMR studies, dUMP binding alone does not induce the closure of the protein, and

the pyrimidine ring of dUMP is located away from the catalytic Cys146, which prevents the initiation of the reaction in the absence of mTHF. After bonding with the cofactor, the protein goes into a closed conformation (Figure 1), steering the pyrimidine ring of dUMP into the position where thiolation by Cys146 can occur, and the cofactor to align for reaction with dUMP (Figure 2) [Montfort and Weichsel, 1997].

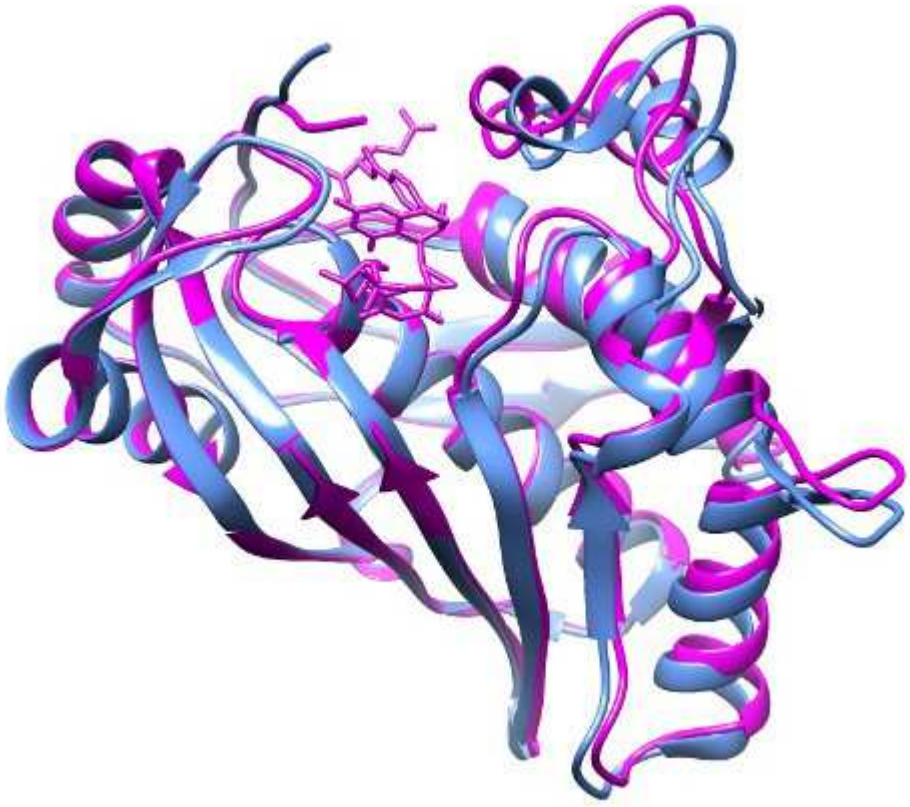

Figure 1 Monomeric unit of *E. coli* TS in open (blue) and closed (magenta) conformation. The blue structure is of an apoenzyme (PDB code 1FTQ), and the closed ternary complex of TS with mTHF and FdUMP(PDB code 1TSN) is coloured magenta. In the upper part of the drawing the closure of protein structure around the ligands is visible.

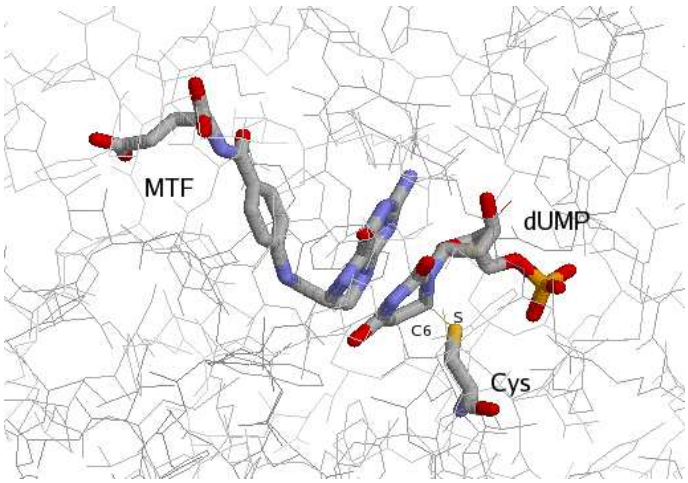

Figure 2 Active centre of TS: ligand arrangement during the ternary complex formation

1.2.2 Thymidylate synthase reaction

TS catalyses the exchange of 5H hydrogen atom of dUMP for a methylene group provided by $N^{5,10}$-methylenetetrahydrofolate (mTHF), with tetrahydrofolate (THF) acting subsequently as a reductant in the reaction and undergoing transformation into dihydrofolate (DHF). Regeneration of mTHF is catalysed by dihydrofolate reductase and serine hydroxymethyltransferase.

The catalysis is initiated with a nucleophilic attack of the active site cysteine (Cys146 in *E. coli* TS) at C6 of dUMP and opening of the methylene bridge in mTHF. A nucleophilic attack of pyrimidine C5 at the folyl iminium ion follows and a covalently linked complex of Cys146, dUMP and the cofactor is formed. Finally, the abstraction of pyrimidine 5H, THF elimination, hydride transfer and dTMP formation conclude the reaction (Figure 3). Some aspects of the enzymatic reaction steps still remain unclear [Montfort and Weichsel, 1997].

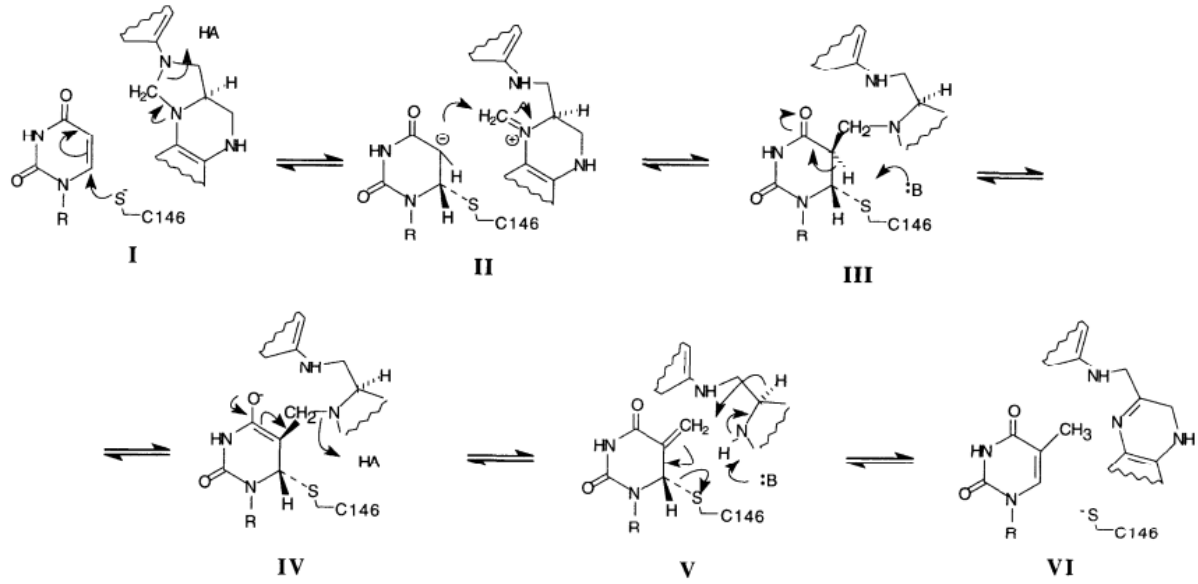

Figure 3 TS-catalysed reaction mechanism. From: [Montfort and Weichsel, 1997]

1.2.3 Dynamical behaviour of thymidylate synthase

The TS ligands are first recruited to a highly conserved large cavity, which progressively closes down throughout the first half of the enzymatic reaction. dUMP forms 13 hydrogen bonds during its initial binding to the apoenzyme and these bonds remain in place throughout the subsequent structural transition. This conformational change is not a domain shift, but rather the concerted movement of several protein chain segments toward the active centre against a stable protein core, comprised largely of the dimer interface. The closure of the active site leads to the alignment of the reactants and to the formation of the majority of contacts between the cofactor and protein, including hydrophobic interactions of the cofactor PABA and pterin rings with the protein side chains and a hydrogen bond network between the C-terminus and pterin ring. It is

suggested that the first chemical step in the catalytic reaction is facilitated by the precise alignment of the cofactor against dUMP. This alignment depends on the conformational change upon cofactor binding, contributing to catalysis by positioning a general acid Glu-58 (according to *E. coli* TS numbering scheme) near the methyl transfer site. It also brings together dUMP and three highly conserved residues, Asn-177, Glu-58, and His-147 (*E. coli* TS numbering scheme) into a new hydrogen bond network (see Figure 13). It is also postulated that the conformational change may contribute to catalysis through conformational strain of the substrate and cofactor that leads to their activation, i.e. opening of the imidazolidine ring of the cofactor to produce the iminium ion intermediate that participates in the covalent ternary complex. The ligands in closed ternary complexes are also forced closer to the catalytic cysteine. After the covalent ternary complex is formed, only minor changes to the protein conformation take place until the products are released. The ligands' binding to one monomer induces relatively small conformational changes in the other protomer that nevertheless inactivates the second binding site (half-of-the-sites reactivity) [Stroud and Finer-Moore, 2003].

Human thymidylate synthase, unlike the enzyme from other species, exists in the equilibrium of active and inactive conformations. In the inactive conformation, the catalytic cysteine is twisted almost 180° so that it points toward the dimer interface and is unable to bind dUMP without a major conformational change in the protein. It is also postulated that the inactive form of hTS interacts with its own mRNA, thus repressing TS synthesis. The active-inactive transition of hTS involves rearranging of the active-site loop that comprises residues 181-197 (hTS numbering scheme) and includes catalytic cysteine (C195 in hTS) as well as R185 arginine, which is unique for hTS. Computational simulations (targeted molecular dynamics) made by Salo-Ahen and Wade indicate that the active site loop rotation occurs sequentially, ending with C195 positioning in the active site, making the enzymatic reaction possible. It is suggested that the hydrophobic pocket unique to the inactive conformation, formed by residues F137, Q138, F142, G143, W182 and L187, might accept ligands that stabilise the inactive conformation without interfering with mRNA binding. Therefore it may be possible to design non-competitive inhibitors targeting this pocket. The advantage of these allosteric inhibitors over orthosteric ones would be lack of the induced drug resistance [Salo-Ahen and Wade, 2011].

1.2.4 Thymidylate synthase ligands

- **dUMP**

Deoxyuridine monophosphate (Figure 4) is a deoxyribonucleotide intermediate in synthesis of deoxythymidine triphosphate (dTTP). It is produced from uridine diphosphate (UDP) which is converted by ribonucleotide reductase to deoxyuridine diphosphate (dUDP), phosphorylated to dUTP, and then dephosphorylated to dUMP by dUTPase. dUMP is subsequently methylated to dTMP by the enzyme thymidylate synthase.

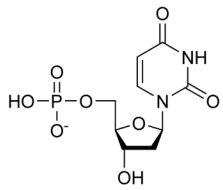

Figure 4 Deoxyuridine monophosphate

- **mTHF**

Folate (known as vitamin B_9) is essential to cell growth and, in mammals, has to be supplied through the diet or by symbiotic gastrointestinal microorganisms. Folate derivatives are needed for DNA synthesis, repairs and methylation, and as cofactors for various enzymes. Natural folates are present in various foods, especially dark green leafy vegetables. A synthetic oxidised form of folate, folic acid, is converted to 7,8-dihydrofolate by dihydrofolate reductase (DHFR). Folic acid supplementation is generally recommended to pregnant women in order to prevent neural tube defects in growing embryos. However, excessive (over 1 mg) daily uptake of folic acid in human may lead to DHFR saturation, excretion of high levels of unmetabolised folic acid, and some degree of inhibition of the reduction of DHFR physiological substrate 7,8-dihydrofolate [Bailey and Ayling, 2009].

Dihydrofolate is further converted by DHFR to 5,6,7,8-tetrahydrofolate (THF), an universal one-carbon fragments carrier in various biosyntheses. THF accepts one-carbon groups (e.g. from serine) and these one-carbon derivatives adopt different oxidation states, from the reduced N5-methyltetrahydrofolate, via N5,N10-methylenetetrahydrofolate (mTHF – the cofactor of TS, Figure 5) and N5,N10-methenyltetrahydrofolate, to the oxidised form N10-formyltetrahydrofolate.

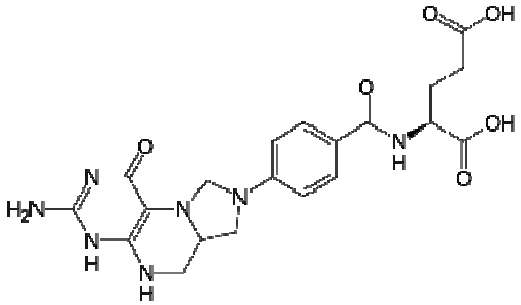

Figure 5 5,10-methylenetetrahydrofolate

In solution, methylenetetrahydrofolate adopts flat conformation with almost co-planar PABA, imidazolidine and pterin rings. However, as this conformation would not fit in the active site of thymidylate synthase, in known ternary complexes the cofactor or its analogues display an L-shaped conformation (Figure 6, see also Figure 2).

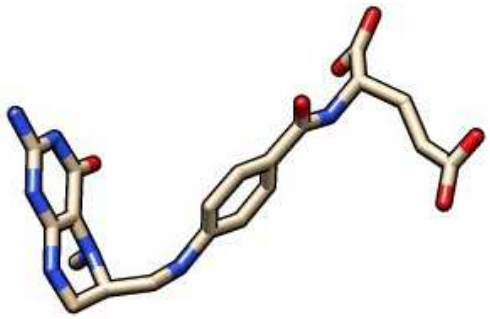

Figure 6 5,10-methylenetetrahydrofolate in bent conformation

As suggested by the results of molecular modelling, the cofactor in its flat conformation binds to thymidylate synthase in an open conformation, and upon closing of the protein the cofactor adopts the higher-energy bent conformation that is closer to the transition state one, held by extensive hydrogen bond contact network. The ordered sequence of ligand binding to TS – first dUMP, then mTHF – becomes random, if methylenetetrahydrofolate acquires a γ-linked polyglutamate tail that independently binds the protein in the absence of dUMP [Montfort and Weichsel, 1997].

1.2.5 Existing drugs targeting TS

Thymidylate synthase is a good target for rational, structure-based drug design. Analogues of both the substrate and the cofactor are used in clinical practice as TS inhibitors. However, substrate analogues may induce drug resistance. One of the mechanisms of this resistance is supposedly the disruption of the autoregulation of translational repression. The TS protein binds to its own mRNA, generating a regulatory feedback loop for translational repression. Only apo-TS can bind the mRNA, ligand binding prevents the interaction. Thus, the presence of inhibitors bound to TS induces overexpression of the enzyme, which negates the therapeutic effect of the drugs [Garg *et al.*, 2010].

In addition to this, many of the TS inhibitors also affect healthy proliferating tissue, leading to a variety of toxic effects caused by damage to bone marrow cells, blood cells and intestinal lining.

- **Substrate analogues**
  - **5-Fluorouracil and other FdUMP prodrugs**

5-Fluorouracil (5-FU, Figure 7), an analogue of uracil with fluorine atom replacing hydrogen at the C-5 position, was one of the first rationally designed antimetabolites, synthesised and patented in the US in late 1950s by Dr. Charles Heidelberger. Thymidylate synthase was not known at that time, but it was suspected that 5-FU, being a close analogue of uracil, may disrupt the biosynthesis of RNA and DNA.

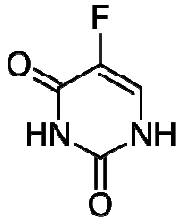

Figure 7 5-Fluorouracil

5-FU enters the cell rapidly via the same transport mechanism as uracil and is converted to a number of active forms, including 5-fluorodeoxyuridine monophosphate (FdUMP), which forms a stable ternary complex with thymidylate synthase and $N^{5,10}$-methylenetetrahydrofolate, thus preventing thymidine synthesis in the cell. The complex is stabilised by high intracellular concentration of reduced cofactor folate, CH_2THF. 5-FU is also converted to fluorouridine triphosphate (FUTP), which is misincorporated into RNA, disrupting its function, and to fluorodeoxyuridine triphosphate (FdUTP), leading to DNA damage, and in consequence to cell death.

5-FU, given intravenously, is still widely used in the treatment of various cancers such as colorectal cancer, breast cancer, and cancers of the aerodigestive tract. The greatest impact of this therapy is observed in colorectal cancer, particularly when 5-FU is combined with newer chemotherapeutics such as irinotecan and oxaliplatin [Longley *et al.*, 2003].

However, 5-fluorouracil is toxic to the patients, causing diarrhoea, and, as mentioned above, induces drug resistance [Mader *et al.*, 1998]. It also undergoes rapid degradation by liver enzymes, which limits its oral bioavailability. Despite the usefulness of 5-fluorouracil and some improvements such as combination therapies, new strategies of using 5-FU are much needed.

One such approach are the modulation strategies intended either to decrease degradation of 5-FU, increase its activation, or increase TS binding activity of FdUMP. One example of a drug successfully used to modulate the activity of 5-FU is methotrexate, an antifolate inhibitor of dihydrofolate reductase (DHFR) which catalyses the conversion of dihydrofolate (DHF) to tetrahydrofolate (THF), the precursor of TS cofactor CH_2THF and a substrate required in purine biosynthesis. The inhibition of purine biosynthesis by methotrexate leads to the increase of the levels of phosphoribosyl pyrophosphate, the cofactor needed for the conversion of 5-FU to FUMP. Thus, when administered before 5-FU, methotrexate enhances its antitumour activity.

The studies on leucovorin (5'-formyltetrahydrofolate) and interferons, despite promising *in vitro* results, revealed no improvement in overall survival of the patients with advanced colorectal cancer; however, using 5-FU/leucovorin combination with oxaliplatin or irinodecan proved more successful.

Other factor that has to be taken into account in predicting response to 5-FU is the individual sensitivity to the drug, determined by genetic factors. Investigating the molecular determinants of chemosensitivity will enable the individualisation of treatment for patients according to their molecular phenotype [Longley *et al.*, 2003].

5-Fluorouracil impacts healthy cells as well as tumours, hence the need for strategies enabling the delivery and activation of the drug only where necessary. An example of a 5-FU prodrug with more specificity for tumour cells is capecitabine (N^4-pentyloxycarbonyl-5′-deoxy-5-fluorocytidine), designed by Miwa *et al.* and approved by FDA in 1998. It is selectively converted to 5-FU in tumour cells by a cascade of three enzymes, one of which, thymidine phosphorylase, is more active in tumour cells than in normal tissues. Capecitabine is characterised by better oral bioavailability and reduced incidence of diarrhoea, although it may cause other adverse side effects such as hand-and-foot syndrome. Other approaches employ enzymes delivered to the surface of tumorous tissues that later activate 5-FU prodrugs locally at the tumour mass [Garg *et al.*, 2010].

The formation of the stable ternary complex with FdUMP and the crystallisation of the complex is the basis of the most of our knowledge on TS mechanism [Hyatt et al., 1997].

- **Analogues of methylenetetrahydrofolate**

  - **Raltitrexed (Tomudex, TDX, ZD 1694)**

The drug raltitrexed (Figure 8), an antifolate inhibitor of thymidylate synthase, is used intravenously since 1998 in the treatment of advanced colorectal cancer in cases where 5-fluorouracil is not appropriate, e.g. due to intolerance or cardiotoxicity. It concentrates easily in the cell and is active in various solid tumours such as gastric cancer, head and neck cancer, malignant mesothelioma, pancreatic cancer, and colorectal cancer. The use of raltitrexed in advanced colorectal cancer is simpler and is associated with less complications than in the case of 5-FU, while leading to an equivalent overall survival [Liu *et al*., 2013]. Raltitrexed may also be suitable for patients with a cardiac history [Kelly *et al.*, 2013].

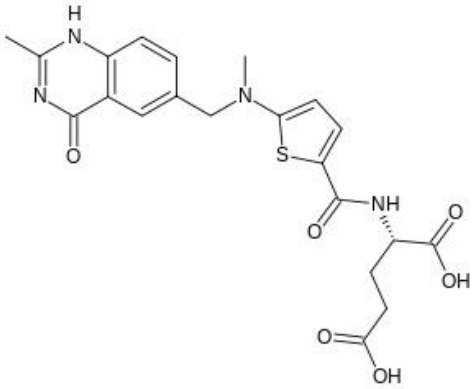

Figure 8 Raltitrexed

  - **Pemetrexed (brand name Alimta)**

Used in the treatment of pleural mesothelioma as well as non-small cell lung cancer, pemetrexed (Figure 9) is a folate analogue that inhibits several enzymes: thymidylate synthase, dihydrofolate reductase, and glycinamide ribonucleotide formyltransferase. Thus it has therapeutic advantage over other, more specific antifolates [DrugBank].

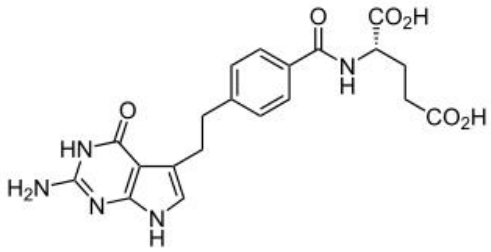

Figure 9 Pemetrexed

- **Allosteric inhibitors**

Thymidylate synthase, as an obligate homodimer, may also be a target for inhibitors disturbing the interaction between protein subunits. Since the use of active-site inhibitors such as 5-FU, raltitrexed or pemetrexed leads to an overexpression of hTS, supposedly due to the loss of RNA regulatory capacity, it is important to find new hTS inhibitors acting through new mechanisms that have no impact on RNA regulation and do not lead to an increase in protein levels. Such allosteric inhibitors are investigated by various researchers. Some diphosphonic acids have been proposed in this capacity, but their potential is limited due to the lack of cellular activity, and in addition to this, the mechanism of action remains unknown. Cardinale *et al*. described the discovery of several peptides, with sequences taken from the hTS dimer interface, that inhibit thymidylate synthase through binding to a previously undescribed allosteric binding site at the dimer interface of the bi-inactive form of hTS. The mechanism of this inhibition involves stabilisation of an inactive form of the enzyme. Administered to ovarian cancer cells, these peptides, unlike the abovementioned drugs targeting the active site, inhibit intracellular hTS – and cell growth – without leading to overexpression of the protein. Since the usage of peptides as drugs is limited due to their degradation by proteases, poor membrane permeation and absorption properties, the subsequent steps will require finding analogous compounds (peptidomimetics) and a detailed analysis of their mechanism of action [Cardinale *et al.*, 2011].

2 The goals of this work

This study is focused on applications of modern molecular modelling techniques, such as MD simulations and molecular docking, into the pre-clinical stages of drug design and development process. The molecular target chosen in this work was the protein thymidylate synthase (TS) – an enzyme playing a pivotal role in cancer invasion, progression, and metastasis. This makes TS a great target for development of novel oncologic drugs. Thus, detailed studies of picosecond-to-nanosecond time scale dynamics of this enzyme with its known ligands were performed and the observed binding modes were investigated using molecular dynamics simulations. MD simulations also enabled the analysis of the energetics of enzyme-inhibitor complex formation.

The results of this part were required for setting up and interpreting the results of virtual screening of the databases of small organic compounds in search of potential new inhibitors. The results of this virtual screening focused my attention on compounds with peptoid and mixed peptide-peptoid scaffolds. Combined with the indications that the

competitive (orthosteric) inhibitors may be of limited usage in clinical practice, this focus on peptoid compounds gave rise to the next part of this project: the design of allosteric TS inhibitors with the peptoid scaffold.
In addition to using existing methodologies in these processes, the last part of the work presented in this thesis included the development of new tools for structure-based ligand design – namely, libraries of peptoid residues in the format ready for molecular simulations. Moreover, results of collaboration with teams of experimental researchers working on thymidylate synthase inhibitors and related compounds are presented in the form of attached published papers with short summaries thereof.

Overall, the goals of this work may be summarised as follows:

- The investigation of ligands binding to thymidylate synthase in the active site – dUMP and its analogues, in the presence of cofactor methylenetetrahydrofolate. The underlying idea is to find competitive TS inhibitors that exhibit better long-term behaviour than 5-fluorouracil.
- Studies of thymidylate synthase dynamics, conformational changes and behaviour in the presence of various ligands in the active site.
- The design of new thymidylate synthase inhibitors, both binding to the active site (dUMP analogues) and to other sites at the surface of the protein (allosteric inhibitors, e.g. peptidomimetics).

3 Methods

Drug discovery requires a vast investment of time and resources. Although clinical trials are undoubtedly the most expensive part of that process, significant costs are also associated with preclinical phases and the hit rates in conventional screenings are rather low: it is estimated that only five out of 40 000 compounds tested in animals reach clinical trials [Kapetanovic, 2008]. Therefore, approaches alternative to conventional screening are being sought in order to facilitate the drug development and optimise the resources in terms of time, manpower, and costs. In the past two decades, vast development of *in silico* tools applied to preclinical drug development was made possible due to the increase in computational power, software development, and constantly growing amount of publicly available data such as solved crystal structures of important protein targets. Computational approaches are currently used in virtual screening, hit detection (identification of drug candidates), hit-to-lead conversion, lead optimisation (reducing the side effects and improving ADMET properties of the selected compounds), and in the target assessment (target dynamics, binding selectivity/off-target binding). These are done primarily by molecular docking methods and molecular dynamics (MD) simulations.

Molecular docking, in principle, predicts the optimal binding pose (geometry and orientation) and its energetics (and therefore binding affinity) of a drug candidate upon binding to its cognate receptor. It can thus be used to rank the list of drug candidates according to their predicted binding energy (virtual screening) in order to select the most promising candidates for further development, that is, possible modifications and

experimental testing. The use of molecular docking techniques typically leads to a relatively short list of predicted binding poses, amongst which a near-native binding mode is often observed [Ritchie, 2008].

Recent results from the CAPRI (*Critical Assessment of PRedicted Interactions*, a community-wide experiment set up to monitor progress in the field of molecular docking) blind docking tests show constant improvement in the currently used docking algorithms, including automatic docking servers [Lensink and Wodak, 2013]. Numerous studies described in details the types of algorithms and scoring functions commonly used in molecular docking protocols, their accuracy, robustness, approximations used, and their advantages and limitations [Cheng *et al.*, 2009, Li *et al.*, 2014 a and b]. In general, regardless of their type (force field-based, empirical, knowledge-based, hybrid, etc.), the scoring functions implemented tend to cover reliable models of electrostatics, van der Waals, and solvation effects, and therefore are very useful in drug design applications.

Methods based on molecular dynamics (MD) simulations provide detailed insight into the nature of potential drug-target interactions by representing the interacting species as a conformational ensemble that follows the laws of statistical thermodynamics. As such, MD simulations are very valuable in the assessment of the dynamics of such complexes in short time scales (typically tens to thousands of nanoseconds, occasionally several microseconds). The analysis of MD trajectories allows for monitoring of the conformational changes induced by binding of a drug molecule to its molecular target, tracking changes in the hydrogen-bonding network, studying the role of the solvation effects in drug-target binding, and assessing the stability and energetics, including enthalpic as well as entropic contributions. Trajectories of MD simulations and their post-processing (e.g. MM-PBSA method, see Section 3.2.3) enables detailed and accurate analysis of binding of very closely related ligands and for rational direction of development of better, more tight-binding compounds.

3.1 Molecular docking

Molecular docking is a method of quickly predicting the location, orientation, and conformation (the binding pose) of one molecule (e.g. a small ligand – drug candidate) versus another (e.g. a protein). Fast, crude docking algorithm applied to a set of putative ligands may serve as a high-throughput tool of hit identification (see Section 1.1). It is also useful during lead optimisation stage, allowing for the testing of various modifications of the lead structure before synthesis.

The aim of docking is basically to find a correct structure of the complex, e.g. of an enzyme (E) and inhibitor (I):

$[EI]_{aq} \rightleftharpoons [E]_{aq} + [I]_{aq}$

The free energy of binding $\Delta G = -RT\ln K_A$ is related to binding affinity K_A and inhibition constant K_i by the equation:

$$K_A = K_i^{-1} = \frac{[EI]}{[E][I]}$$

For the purposes of high-throughput docking and preliminary scoring of docked compounds, binding energy can be crudely estimated as a sum of Coulombic (E_{coul}) and van der Waals (E_{vdW}) interactions:

$$E_{coul}(r) = \sum_{i=1}^{N_E} \sum_{j=1}^{N_I} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

where N is the number of atoms in molecules E and I, and q is the atomic charge on each atom;

$$E_{vdW}(r) = \sum_{j=1}^{N} \sum_{i=1}^{N} 4\varepsilon \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right]$$

where ε is the well depth of the potential and σ is the collision diameter of the atoms i and j, respectively.

The process of docking a ligand to a receptor involves:

1. Posing – finding orientations and conformations, within a given degree of flexibility for each molecule, fitting into the active site of the receptor. A balance must be found between the number of degrees of freedom of the molecule – with the aim to find the right conformation – and the algorithm speed that would permit the evaluation of thousands of molecules in a short time. There are several methods of treating flexibility of both the receptor and the ligand. For example, the ligands may be divided into fragments that are docked into the active site and linked again; or a “rigid core” of the molecule may be docked first and flexible “side chains” added later, one molecular fragment at a time. A set of pre-generated conformations can also be used. Other approaches involve stochastic methods of conformational space exploration, or the use of molecular dynamics (MD) simulations. Because of the approximations used and the requirement for the docking algorithm to be fast, the receptor flexibility is usually somewhat compromised. Amino acid side chain rotamer libraries and protein ensemble grids are used along with the algorithms removing the less favourable conformations and steric clashes. The initial pose score can be calculated as an approximate measure of how the ligand fits into the binding pocket of the receptor.
2. Ranking – evaluation (scoring) of the “best” predicted ligand poses in order to indicate which ligands are suitable for further investigation, usually by approximating the free energy of binding (the so-called binding score or the energy score). The simplest scoring method is to estimate binding energy as the sum of Coulombic and van der Waals components given above, and in some cases to normalise the resulting energy by dividing it by number of heavy atoms in the molecule – since the larger the molecule is, the more interactions it has with the receptor. Other scoring approaches may involve the use of force fields, empirical scoring functions, knowledge-based scoring functions, and other

schemes. Entropic and solvation terms may also be taken into account [Kitchen *et al.*, 2004].

After the best ligands for further development – lead compounds – have been selected, lead optimisation phase (see Section 1.1) may begin with the use of further docking procedures for combinatorially modified leads (Section 4.2.4), molecular dynamics simulations, and *in vitro* tests.

3.2 Molecular dynamics

3.2.1 Basic concepts of molecular dynamics

The basic concept of molecular dynamics is to simulate time-dependent behaviour of a system, e.g. a protein, protein-ligand complex, or macromolecular complex. In general, atoms are treated as points in space with charges assigned and may or may not be connected with bonds. The movements of the atoms are described by numerically solving Newton's equations of motion.

The force field V describes electronic energy as a function of nuclei configuration $\mathbf{R}$. The force field gradient gives the forces acting on the atoms. In classic molecular mechanics (MM) and molecular dynamics (MD), the force field is approximated by the following function:

$$V(\mathbf{R}) = \sum_{bonds} k_r (r - r_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} k_\phi [1 + \cos(n\phi + \phi_0)] +$$

$$+ \sum_{atom\_i} \sum_{j \neq i} 4\varepsilon_{i,j} \left[\left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{6} \right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\varepsilon_0 r_{i,j}}$$

The first term of the equation above describes chemical bonds within the harmonic model in relation to a certain reference length for a given pair of atoms r_0; the second term describes bond angles in the same way in relation to a reference angle θ_0; the third term pertains to torsional interactions; the fourth term describes Van der Waals forces between non-bonded atoms; and the last term gives Coulombic interactions between atomic charges that are assumed to be fixed. In the classic MM/MD models used in this work the bonds are "unbreakable" and no new bonds are created.

Molecular mechanics is a method of finding a stable configuration of the nuclei (called equilibrium geometry) by minimising $V(\mathbf{R})$ with respect to the nuclear coordinates. The resulting geometry, being a local minimum of potential energy, is usually close to the starting one, i.e. for a reasonable starting geometry no significant conformational changes occur.

In a molecular dynamics simulation time is taken into account as well as the temperature and acceleration of the atoms.

The method is based on Newton's equation of motion $\mathbf{F}_i = m_i \mathbf{a}_i$, or:

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2}$$

where V is the potential energy of the system. In order to obtain positions of the atoms r_i after the time t, the equations of motion must be solved numerically. The integration algorithms are based on the Taylor series expansion:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \frac{d\mathbf{r}(t)}{dt}\delta t + \frac{1}{2}\frac{d^2\mathbf{r}(t)}{dt^2}(\delta t)^2 + ... = \mathbf{r}(t) + \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)(\delta t)^2 + ...$$

$$\mathbf{v}(t+\delta t) = \mathbf{v}(t) + a(t)\delta t + ...$$

The initial positions are assumed (e.g. according to experimental results or other simulations), and the initial velocities are assigned randomly according to Maxwell-Boltzmann distribution for a given temperature. Then the positions and velocities of atoms after one time step are calculated e.g. using the Verlet algorithm, assuming linear change of velocity and acceleration during the time interval δt. In this way conformational space of the system can be explored [Piela, 2014].

A typical molecular dynamics simulation of a protein-ligand system is performed in constant pressure and temperature (NPT) conditions, and involves:

- choosing initial coordinates of the system
- energy minimisation of the structure in order to remove any steric clashes and bad electrostatic contacts that might occur after the initial construction of the structure
- adding solvent (if explicit solvent molecules are to be used)
- adding counter-ions, in order to maintain the neutral charge of the whole system
- heating phase: the simulations starts at low temperature that is gradually, during a short MD run, increased to the target temperature (e.g. 300 K)
- equilibration phase: achieves the correct density and stabilises the properties of the system with respect to time
- actual simulation – production runs, typically of nanosecond time scale, up to microseconds. Production runs may comprise of a single trajectory or multiple ones (starting from the same, equilibrated structure, in order to improve the sampling and the statistics).

During a MD run, the velocities of the atoms are regularly checked and scaled in order to maintain the desired temperature. Periodic boundary conditions are often used.

The molecular dynamics simulations are useful in:

- evaluation of conformational changes of the system (RMSD, principal component analysis)
- calculating free energy and affinity (MM-PBSA)
- assessment of the stability of the system
- calculation of enthalpic and entropic energy components
- comparison to experimental results (e.g. NMR, crystallography)

- generation of conformational ensembles (e.g. for molecular docking).

3.2.2 Preparing the MD simulations

Geometry optimisation

In order to obtain correct energies and atomic charges, an accurate geometry of the molecule – that is, an energy minimum – is needed. In the case of relatively small molecules such as the ligands of thymidylate synthase, one can search the potential energy surface with quantum mechanical methods. In the work described here, geometry optimisation was performed using Berny algorithm implemented in Gaussian 03/09 program suite and described in the Opt section of Gaussian 09 User's Reference [G09 User's Reference], at the HF/6-31G* level, sometimes with preliminary optimisation using DFT (B3LYP) to get better starting geometry for the more demanding *ab initio* computations. This procedure yields local minima, thus the starting structures resembling the ones present in X-ray data were used.

Since a protein is too big a system for *ab initio* quantum mechanical geometry optimisation, starting structures of the complexes were optimised with molecular mechanics only, starting with X-ray structures for the protein and QM-optimised geometries for the ligands. The initial positions of the ligands versus the protein were assumed according to the geometries of analogous complexes (Section 4.2.1) or to the results of molecular docking (see Sections 3.1 and 4.2.2).

Restrained Electrostatic Potential (RESP) method of determining atomic charges

All molecular mechanics and molecular dynamics calculations rely on the concept of atomic charges. However, there is no such thing as "true" atomic charge, one can only speak of the distribution of electrons around the nuclei in the molecule. Some approximate construct must be used in order to derive atomic point charges that allow for accurate computer simulations for the system.

Using *ab initio* calculations, one can derive atomic charges by fitting them so that electrostatic potential (ESP) calculated for a grid of points around the molecule is accurately reproduced. For this, Merz-Singh-Kollman scheme us used on the HF/6-31G* level. This method yields atomic charges that accurately reproduce multipole moments and intermolecular interactions. However, they are basis set dependent and conformation dependent. Moreover, the charges for "buried" atoms (less exposed to solvent, such as an sp^3 carbon) are poorly determined, since the ESP points used to fit the charges lie outside the Van der Waals surface of the molecule. A hyperbolic restraint function is therefore used in the fitting, which reduces the fluctuation in charges of buried atoms without seriously impairing the quality of the overall fit to ESP. Moreover, the charges for equivalent atoms (i.e. atoms rapidly exchanging during molecular dynamics, e.g. hydrogen atoms of methyl groups) should be identical. In the case of methyl groups, a two-stage fit is used – the first stage is performed without forced symmetry and with weak restraints, and the second stage refit concerns only methyl groups, with forced symmetry and strong restraints.

The charges derived with the use of the RESP method are suitable for quantitative and qualitative description of a molecule. RESP charges are consistent with chemical intuition, exhibiting moderate fluctuations between related functional groups, and still reflecting the local chemical environment [Bayly *et al.*, 1993]. The calculations of conformational energies, hydrogen bonds energies, and free energies of solvation made by Cornell *et al.* show that the two-stage RESP charges reproduce both intermolecular and intramolecular energies and structures quite accurately [Cornell *et al*., 1993].

Amber parameterisation

In Amber and other similar programs for molecular mechanics / molecular dynamics calculations, force field parameters are specified in a file containing values of r_0, θ_0, and φ_0 for any combination of atom types that is likely to appear in a given calculation. The assignment of an atom type depends on the hybridisation state of an atom (e.g. sp^2 or sp^3) and its chemical environments (e.g. sp^2 nitrogen in amide groups). In the calculations described here, PARM99 and GAFF force field parameters were used, with modifications in the cases where less typical bond angles were present in the ligand molecule and thus the force field file lacked the corresponding angle and dihedral parameters. In these cases parameters for similar atom types were used.

3.2.3 Analysis of the MD simulation

Visualisation of the results

The results of the molecular dynamics calculation can be viewed frame by frame as a film, or individual frames or averages can be analysed visually, in order to spot any irregularities, and to get an idea on whether, and to what degree, does the overall conformation of the system change. In this work, visualisation was performed and all the drawings depicting molecular conformations were made with the use of freeware molecular graphics programmes: UCSF Chimera [Pettersen *et al.*, 2004], RasMol [Bernstein, 2000], VMD [Humphrey *et al.*, 1996], and Molden [Schaftenaar and Noordik, 2000]. These programmes, along with XleaP (part of the Amber package), were also essential in preparing all structures for simulations and for visual inspection of the results obtained.

RMSD

The root-mean-square deviation (RMSd) is the measure of the average distance between the atoms (all of them, protein backbone only, or selected groups) of the superimposed structures. In the analysis of a molecular dynamics run the RMSd is a tool to estimate quantitatively the extent of conformational changes in time and the stability of the system.

It is calculated according to the formula:

$$RMSd = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta_i^2}$$

where δ is the distance between N pairs of equivalent atoms.

End-state free energy calculations: MM-PBSA

The binding free energy in a noncovalently bound receptor-ligand complex may be calculated by subtracting the free energies of the nonbound receptor (apo-protein) and ligand from the free energy of the complex:

$$\Delta G_{binding,solvated} = \Delta G_{complex,solvated} - [\Delta G_{receptor,solvated} + \Delta G_{ligand,solvated}]$$

The terms $\Delta G_{solvated}$ on the right-hand side of the above equation are estimated as:

$$\Delta G_{solvated} = E_{gas} + \Delta G_{solvation} - TS_{solute}$$

where E_{gas} denotes gas-phase energies, often calculated by molecular mechanics from the force field, $\Delta G_{solvation}$ is the solvation free energy calculated using an implicit solvent model (e.g. Generalized Born), and ΔS – the entropic contribution – is estimated e. g. by normal mode analysis. In practice, in the end-state calculations these energies are estimated from the average values for an ensemble of representative structures:

$$\Delta G_{solvated} \cong \langle E_{gas} \rangle + \langle \Delta G_{solvation} \rangle - T \langle S_{solute} \rangle$$
$$= \frac{1}{N}\sum_{i=1}^{N} E_{i,gas} + \frac{1}{N}\sum_{i=1}^{N} \Delta G_{i,solvation} - \frac{T}{N}\sum_{i=1}^{N} S_{i,solute}$$

where *N* is the number of analysed frames, *i* – the frame index.

The ensembles can be extracted from a molecular dynamics or Monte Carlo run, either from a single trajectory for the complex, or from trajectories generated separately for each state [Miller *et al.*, 2012].

4 Summary of the results

4.1 Published paper

The initial steps of thymidylate synthase reaction were investigated using QM/MM approach.

It is assumed that the Asn177 (*E. coli* numbering scheme) residue stabilises, by hydrogen bonding, the partial negative charge on O4 of covalently bound dUMP (for the numbering of atoms see Figure 10). The “switch effect” hypothesis developed in this work is that the screening of this partial negative charge facilitates the formation of the C—S bond between C6 of dUMP and Cys146 of the enzyme. In other words, without the partial negative charge on O4 the dUMP-enzyme bond is not stable.

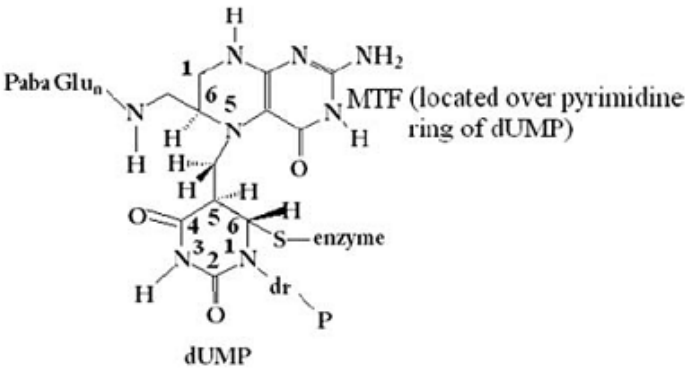

Figure 10 Schematic diagram of the TS-dUMP-MTF complex

In order to verify this hypothesis, three models of TS ternary complexes with methylenetetrahydrofolate (MTF) with dUMP, FdUMP, or dUMP-OH, respectively, were created using the X-ray diffraction structure of *E. coli* TS complexed with dUMP and THF (PDB accession code 1KZI) as the starting point. Models of three different sizes were used; the largest one, comprising of residues in 10A sphere (within 10A from the S atom of Cys146), was investigated by combined quantum mechanical / molecular mechanics (QM/MM) routine using two-layer ONIOM-ME scheme. The internal part, containing residues within 5A sphere, was computed at the B3LYP/LANL2DZ level, and the external layer was studied by MM with Amber force field. The smaller model consisted only of the residues within the 5A sphere and was studied at the B3LYP/LANL2DZ level. For the transition state calculations a small model consisting only of uracil and selected atoms from its environment was used (Figure 11).

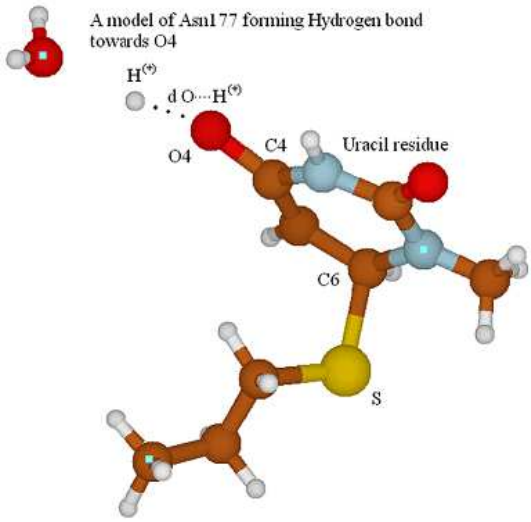

Figure 11 The model of the transition state for the proton movement. H_3O^+ is mimicking the presence of Asn177 and the 6-mercaptopropane group represents Cys146

All calculations were performed with the Gaussian03 software.

It was concluded that the nucleophilic attack of Cys146 on dUMP appears to be concerted with the presence of a proton bound to O4 atom of dUMP, or a proton donor forming hydrogen bond with O4 atom. Moreover, this “proton switch” stabilises or destabilises the sulphur – carbon S(Cys146)—C6(dUMP) bond. The “switch effect” does not occur when there is sulphur or a –SH group in place of the O4 atom.

Additional calculations indicate that the "switch effect" occurs also in C5-substituted analogues of dUMP, i.e. FdUMP, dUMP-OH, dUMP-CH_2OH, and dUMP-OCH_3.

This work was conducted in cooperation with M. Fusaro, A. Jarmuła, A. Leś, and W. Rode [Fusaro *et al*, 2010], and the full text can be found in Appendix 1.

4.2 Non-published work

4.2.1 Hydroxyl analogues of dUMP

Analogues of the substrate

In search of new inhibitors of thymidylate synthase, certain hydroxyl analogues of the substrate, where the 5H hydrogen atom of dUMP was substituted by –OH, –CH_2OH and other related groups, were investigated experimentally by W. Rode's team in the Nencki Institute of Experimental Biology [Dzik *et al.*, 1997].

The above mentioned experiments showed that both dUMP-OH and dUMP-CH_2OH are competitive vs. dUMP inhibitors of thymidylate synthase. The dUMP-OH demonstrates some substrate-like activity in that mTHF is slowly converted into dihydrofolate, however, the products of this reaction remain unknown. The product analogue, dUMP-CH_2OH, acts as a slowly binding inhibitor inactivating the enzyme in a time-dependent matter [Dzik *et al*, 1997]. The computational simulations [Jarmuła *et al.*, 2000] suggested that the difference of behaviour between dUMP-OH and dUMP-CH_2OH is related to different acidity of the respective substituents. In dUMP-OH and dUMP-OCH_3, the C5-O5 bond is partially double due to the resonance with the pyrimidine ring. This is not the case with dUMP-CH_2OH, where the bond between C7-O7 is strictly single. Hence, in dUMP-OH the O5 atom bears positive charge, promoting deprotonation of the 5-hydroxyl group. This computational result is supported by UV absorption spectra of 5-hydroxyuridine and 5-hydroxy-2'-deoxyuridine.

In addition to the above, the presence of an additional hydroxyl group in the chemical environment of the modified uracil ring should promote new hydrogen bonds in the active centre and impact the recognition of the analogue by the enzyme and the course of enzymatic reaction. For comparison, dUMP-OCH_3, a ligand similar to dUMP-CH_2OH, but without additional hydrogen bonding ability, was also taken into account. As reference, identical MD simulations for dUMP and the known TS inhibitor, FdUMP, were also performed. Figure 12 shows hydroxyl analogues of dUMP that were investigated along with dUMP and FdUMP.

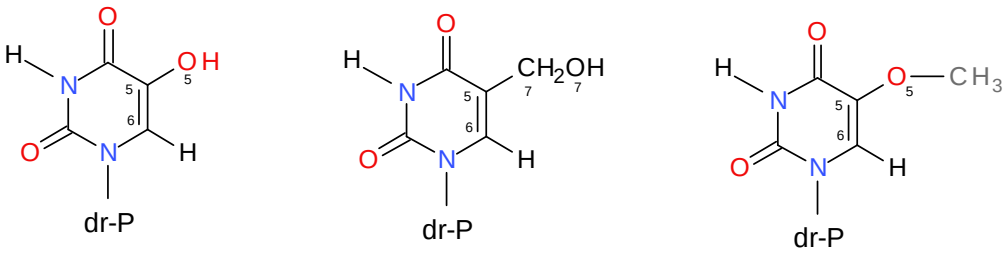

Figure 12 Hydroxyl analogues of dUMP: dUMP-OH, dUMP-CH_2OH, and dUMP-OCH_3; dr-P denotes deoxyribose-phosphate group

Modelling of the interactions in the active centre

The crystal structure of thymidylate synthase used in the simulations was a modified 1TSN PDB structure (*E. coli* thymidylate synthase bound to FdUMP and methylenetetrahydrofolate). The cofactor in this ternary complex is in its bent conformation with the imidazoline ring open, and is covalently bound to FdUMP, which in turn is bound to the enzyme by C6-Cys146S bond. This structure, representing a model of the first stage of the enzymatic reaction, was the starting point for all MD simulations with the five abovementioned TS ligands.

The dynamics of TS active site upon binding to the studied ligands in the presence of the cofactor was investigated in line with the description of the active site in the paper [Finer-Moore *et al.*, 2003] (Figure 13).

MD protocol

The geometries of the ligands and the cofactor were optimised with Gaussian 03 at HF/6-31G* level, then RESP protocol (see Section 3.2.2) was used to calculate charges. Then the structures were parameterised using modified parm99 (for the ligands) and GAFF (for the cofactor) force fields available in the Amber8 package [Amber8], and finally minimised by MM calculation made in Amber8. Then the ligands were inserted into the active site of the protein, aligned in the same position as FdUMP in the 1TSN structure.

Then 20 ns molecular dynamics in periodic boundary conditions and in the presence of explicit TIP3P water in a 12A solvent box was performed with Amber8 *sander* programme, according to the following protocol: 50 ps of heating from 0 to 300 K, 50 ps of equilibration (NPT ensemble), and then 20 nanoseconds of molecular dynamics simulation. The same simulation was performed for the enzyme without any ligands (apo-TS) and for the enzyme bound to the cofactor only, in order to compare the dynamics of the protein, and to find ligand- or cofactor-induced conformational changes within the protein.

Post-MD processing

The MD trajectories were then analysed with *ptraj* module available in AmberTools. Investigated were the binding energies, geometries of the active centre along with hydrogen bonds between the ligands and the enzyme, and protein tertiary structure features.

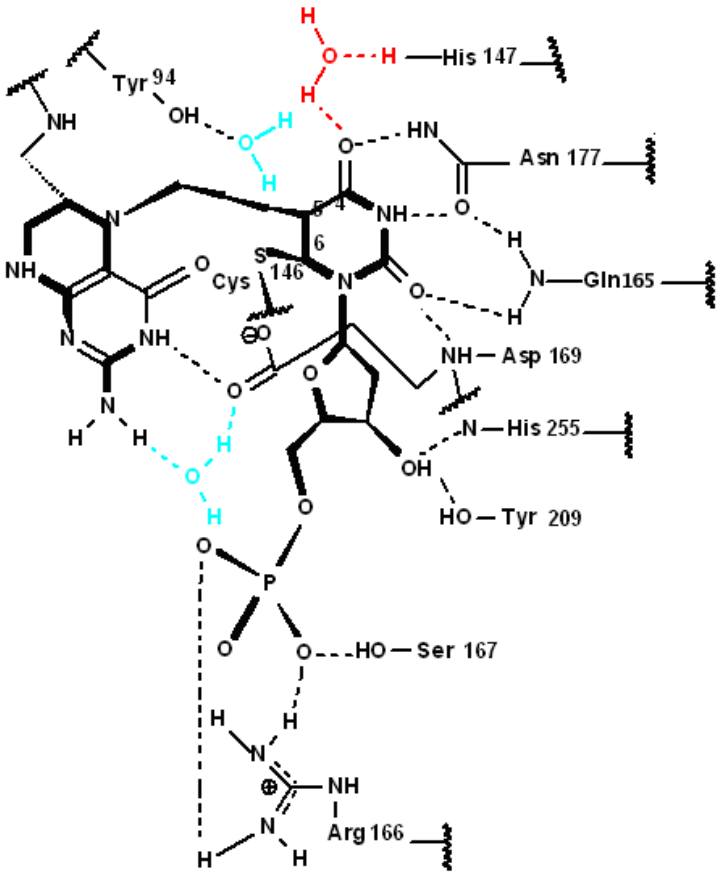

Figure 13 Active centre of TS: residues essential for catalysis (according to: [Finer-Moore *et al.*, 2003], amino acids numbered as in *E. coli* TS)

The following distances between heavy atoms involved in protein – ligand binding, corresponding to hydrogen bonds shown in the above diagram of the active centre, were recorded during 20 ns molecular dynamics simulation: between C6 of X-dUMP and S of catalytic Cys146 (Figure 14 a); O2 of X-DUMP and amino-N of Gln165 (Figure 14 b); O4 of X-dUMP and amide-N of Asn177 (Figure 14 c), O of phosphate group in X-dUMP and hydroxyl-O of Ser 167 (Figure 14 d), O of phosphate group in X-dUMP and amino-N of Arg166 (Figure 14 e), and between ligands: C5 of dUMP and the methylenetetrahydrofolate (Figure 14 f). In the case of the phosphate group of dUMP (Figure 14 d and e), the phosphate oxygen atoms that remain the closest to the respective amino acids are shown (the rotation of phosphate group during the dynamics is visible on the charts as rapid changes of the distance from 2.7 to over 4.5 A). The distances along the dynamics are shown as simple moving averages, where the average distance from last 1000 data points (1 nanosecond) is shown so that the fluctuations are smoothed out and the charts are more readable. The X axis shows the simulation time (in picoseconds), the Y axis shows inter-atomic distances in angstroms.

The following abbreviations are used on the charts:

DMP = dUMP; FMP = FdUMP; MOD = dUMP-OCH_3; OCD = dUMP-CH_2OH; OHD = dUMP-OH

From looking at the series of charts (Figure 14) it is apparent that among the ligands investigated, dUMP-OH (denoted as OHD, light blue line on the charts) is strongly

involved in the interaction with the enzyme via the hydrogen bonds network. Along with dUMP (denoted as DMP, red line), it also remains in the position suitable for binding the catalytic cysteine (which in *E. coli* is Cys146) (Figure 14 a). dUMP-OH also stays most strongly bound by ligand-binding amino acids Asn177, Ser167, and Arg166 (Figure 14 c, d and e, respectively). dUMP-OH also stays approximately in one place with respect to Gln165 (Figure 14 b).

On the other hand, dUMP-CH_2OH (denoted as OCD, purple line) seems to be relatively weakly bound by the active site, only its phosphate group stays in place with respect to Ser167 and Arg166 (Figure 14 d and e). Not surprisingly, dUMP-OCH_3 (denoted as MOD, dark blue line), having a bulky substituent at C-5 that does not take part in hydrogen bonds, stays rather far from the catalytic cysteine and from Gln165, which normally stabilises the ligand on the other side of the pyrimidine ring.

In general, these three investigated ligands behave similarly to the known inhibitor, FdUMP (denoted as FMP, green line) and to the substrate, dUMP (denoted as DMP, red line). The hydrogen bond network in the case of all five compounds remains in place, at least partially. As could be expected, FdUMP exhibits the weakest affinity to methylenetetrahydrofolate (Figure 14 f), while dUMP-OH, which might be stabilised by more hydrogen bonds in the active site, remains relatively close to it, keeping the alignment of the enzymatic reaction.

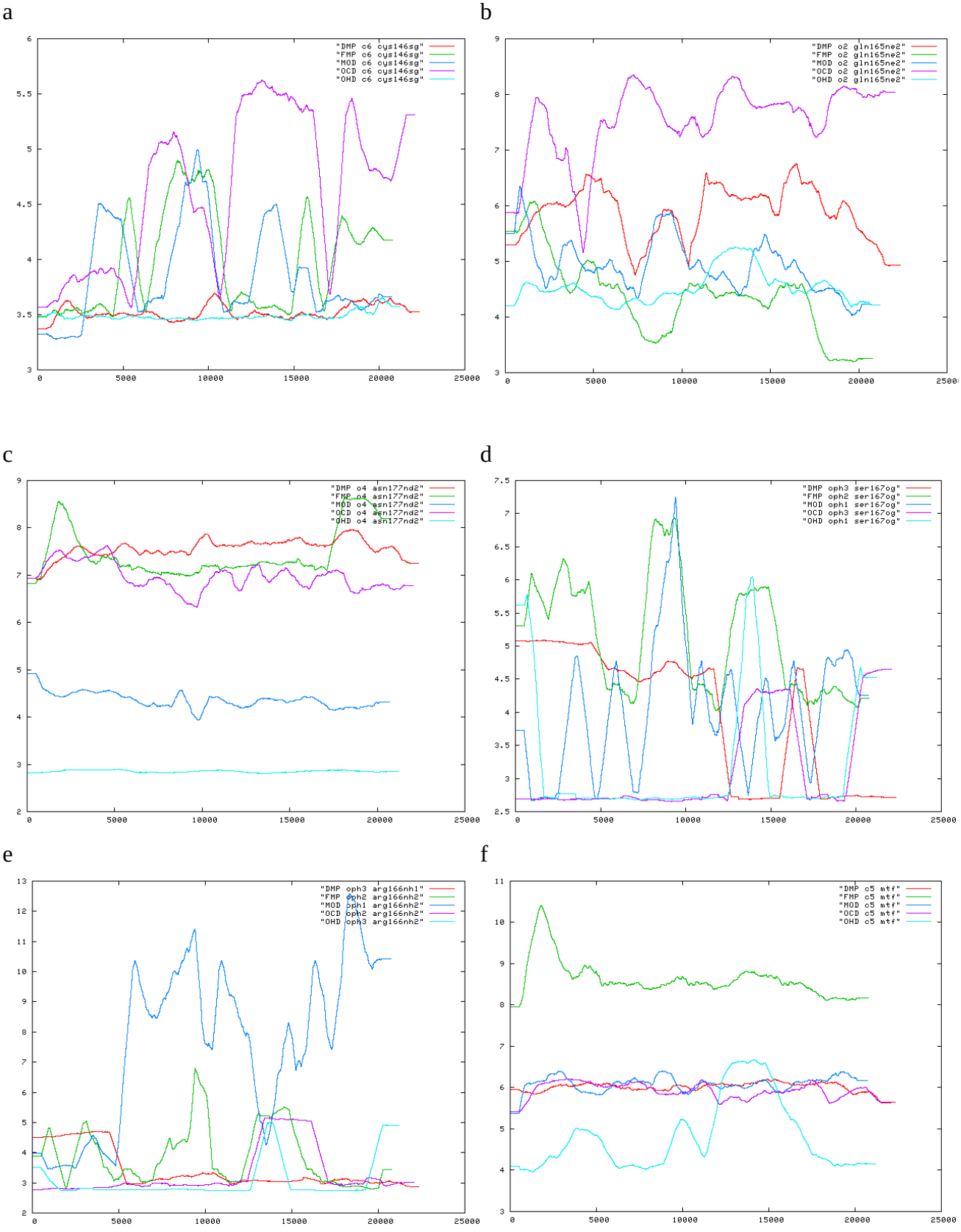

Figure 14 20 ns runs for five TS-ligand complexes, distances measured between: a) X-dUMP_C6 – Cys146_SG, b) X-dUMP_O2 – Gln165_NE2, c) X-dUMP_O4 – Asn177_ND2, d) phosphate oxygen of X-dUMP and Ser167_OG, e) phosphate oxygen of X-dUMP and Ser167_OG, f) X-dUMP_C5 and MTF_C11

MM-PBSA calculation – estimation of binding free energies

For all the complexes, MM-PBSA calculations (see Section 3.2.3) were performed in order to estimate binding energies and compare them to the available experimental data.

The results are presented in Table 1.

Table 1 Free energies of binding for dUMP, FdUMP and hydroxyl analogues of dUMP. Experimental K_i from: [Jarmuła et al.*, 2000]; ΔG_{Exp} calculated according to the formula: $\Delta G_{Exp} = RTlnK_{i,Exp}$*

Ligand	ΔG_{total} (kJ/mol)	$K_{i,Exp}$(microM)	ΔG_{Exp} (kJ/mol)
DMP	-44.03 ± 8.38	3.2	-31.56
FMP	-86.99 ± 10.25	0.02	-44.22
MOD	-35.56 ± 9.01	3.7	-31.20
OCD	-38.77 ± 10.54	3.7	-31.20
OHD	-58.4 ±11.83	3.2	-31.56

4.2.2 Virtual screening and docking

Ligand selection and preparation

In the present study 8060 ligands were selected for virtual screening in the "hit detection" part of the project. For virtual screening, 4890 experimental, purchasable, small-molecular compounds available in ZINC database [Irwin *et al.*, 2012] along with 1491 FDA-approved drugs and 82 FDA-approved nutriceuticals were chosen. In addition to this, 2 sets of natural products: AfroDB (954 compounds, [Ntie-Kang *et al*., 2013]) and NuBBE (643 compounds, a database of natural products found in the environment of Brazil [Valli *et al*., 2013]) were selected and tested.

Partial atomic charges on all tested ligands were assigned using the AM1-BCC methodology [Jakalian *et al.*, 2002] and obtained from the ZINC database.

Molecular docking protocol

In all calculations targeting the catalytic pocket of the thymidylate synthase (TS), the crystal structure of human TS bound to dUMP and Raltitrexed (PDB code: 1HVY) was used as a starting structure of the target [Phan *et al.*, 2001]. This structure is a homodimer, in its closed conformation, and represents the "bioactive" conformation of the enzyme. Prior to the docking procedure, it was modified by removing the covalent bond between the catalytic cysteine C195 and dUMP (which was present in both subunits), deleting ligands, adding hydrogen atoms, and relaxing side chains of all protein residues by MM energy minimisation, 5000 cycles. The binding site for the tested ligands has been selected on one of the monomers. Molecular docking was performed using the UCSF DOCK6.5 suite [Lang *et al.*, 2009], with semi-rigid docking procedure and energy grid scoring in an implicit solvent. The grid spacing was 0.25 A, and the grid included 12 Å beyond the terminal sphere corresponding to the flanking heavy atoms of the Raltitrexed molecule, which was present in crystal structure. The

energy score was the sum of electrostatic and van der Waals contributions. In the course of the docking procedure, each ligand molecule was subjected to 1500 cycles of molecular-mechanical energy minimisation. The number of maximum ligand orientations was 2500. Partial flexibility of the protein during the docking was enabled. After the docking the cluster analysis of the obtained complexes was performed. Complexes with ligand poses more similar (in terms of RMSD on heavy atoms) than 2 Å were assigned to the same cluster. The population of the clusters has been used as one of the criteria of analysing the results. The total number of best-scoring poses analysed was 25.

The docking procedure has been calibrated using crystallised ligands as well as 443 known TS inhibitors obtained from the ZINC database. The set of 386 non-binders was used as a negative control. The procedure successfully reproduced experimentally observed binding modes, and the energy scores for binders were significantly more favourable than for non-binders. These results gave the confidence in the applied methodology.

The putative lead compounds, which were selected by molecular docking, are listed in Table 2 (hits from ZINC database), Table 3 (natural compounds), and Table 4 (competitive inhibitors from the FDA-approved set), respectively.

For the docking studies targeting the allosteric binding site, the crystal structure of human TS dimer bound to a short peptide LSCQLYQR (PDB code: 3N5E) has been chosen as a receptor. The putative ligand binding site has been assigned based on the positions of the heavy atoms of the peptide reported (Figure 15) [Cardinale *et al.*, 2011]. Because of the protein-protein and flexible ligand-protein interactions, the conformational change of protein side chains and loops could occur upon binding of diverse ligands. Thus, 5 conformations of the binding site have been selected and considered independently in the docking studies, and the docking results have been refined using short MD simulations for best-scoring ligands (50 ps). The best scoring leads are presented in Table 5.

Putative competitive (orthosteric) inhibitors

Table 2 Potential orthosteric leads selected from the ZINC database. Starting structure 1HVY, reference score for D16 (Tomudex): -79.22 (X-ray structure) or -78.07 (structure from ZINC)

ZINC ID	Score	No. of heavy atoms	Benign functionality (ZINC)
ZINC38408177	-72.51	35	No
ZINC67541826	-69.71	28	Yes
ZINC14120429	-66.23	27	Yes
ZINC64775540	-63.91	33	Yes
ZINC71703871	-61.74	24	Yes

ZINC ID	Score	No. of heavy atoms	Benign functionality (ZINC)
ZINC08655373	-78.84	35	Yes
ZINC37584283	-71.38	21	Yes
ZINC70311107	-68.13	21	Yes
ZINC37584284	-66.87	21	Yes
ZINC70309711	-65.45	19	Yes
ZINC70311108	-64.23	19	Yes
ZINC50195860	-62.40	20	Yes
ZINC09778504	-61.38	35	Yes
ZINC09778499	-61.13	35	Yes
ZINC70309712	-60.02	19	Yes
ZINC31948501	-60.01	22	Yes

Table 3 Potential leads (orthosteric) selected from AfroDB and NuBBE databases of natural compounds

ZINC ID	Score	No. of heavy atoms	Benign functionality (ZINC)	Name (if available)
ZINC95485963	-60.54	38	Yes	
ZINC28462577	-63.66	40	No	Ochnaflavone
ZINC03871576	-53.17	20	Yes	Apigenin
ZINC84154073	-56.34	37	Yes	
ZINC03874317	-47.63	23	No	Myricetin
ZINC03869685	-46.45	22	No	Quercetin
ZINC84153817	-44.55	24	Yes	
ZINC03872070	-44.31	19	Yes	

Table 4 Potential leads (orthosteric) selected from FDA-approved set of molecules

ZINC ID	Score	No. of heavy atoms	Benign functionality (ZINC)	Name
ZINC03927870	-93.55	24	No	Fludarabine
ZINC03830428	-75.13	35	No	Cefonicide
ZINC12503222	-74.27	20	No	Famotidine
ZINC03833880	-73.45	33	Yes	Rosuvastatin
ZINC01851132	-73.41	31	Yes	Alimta (Pemetrexed)(*)
ZINC03830442	-73.11	36	No	Cefotetan
ZINC01530639	-72.48	30	Yes	Fluvastatin
ZINC29319828	-71.38	30	No	Eprosartan
ZINC03871967	-71.00	27	No	Ceftibuten
ZINC01536934	-69.97	26	No	Ridogrel(**)

(*) Pemetrexed is a known TS inhibitor used in clinical practice

(**) Ridogrel is a dual action drug useful for the prevention of systemic thrombo-embolism and as an adjunctive agent to thrombolytic therapy in acute myocardial infarction. However, there currently are no clinical indications for preferential use of ridogrel over aspirin

Putative allosteric inhibitors

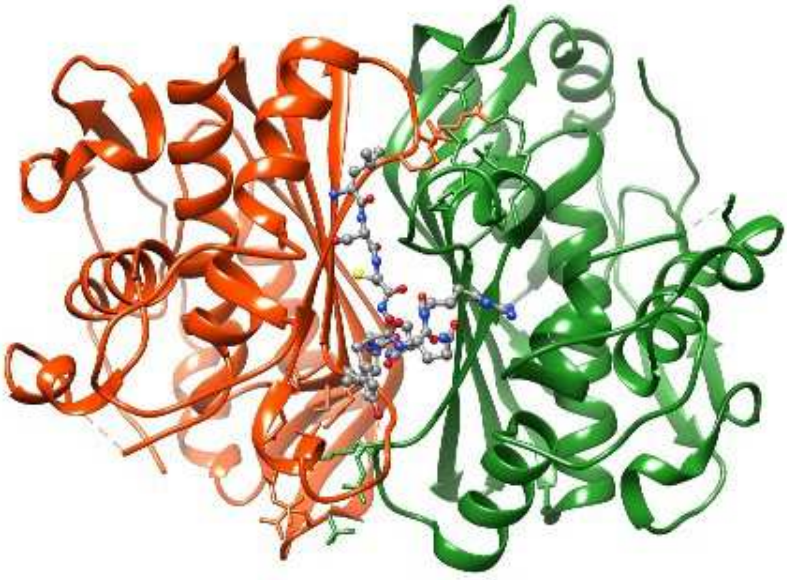

Figure 15 Allosteric binding site at the dimer interface of thymidylate synthase. Binding peptide LSCQLYQR shown in the middle. PDB code 3N5E [Cardinale *et al.*, 2011]

Table 5 Potential leads (allosteric) obtained from ZINC database, peptide and peptoid set. All the molecules are annotated as having benign functionality

No.	ZINC ID	No. of heavy atoms	Score	Remarks
1	**ZINC17911992**	**29**	**-77.44**	**Peptide. Start for peptoid development and for optimisation. Purchasable**
2	ZINC65299894	29	-75.18	Peptide
3	ZINC67661339	29	-75.15	Peptide
4	ZINC62554870	27	-74.99	Peptoid
5	ZINC41398864	29	-74.98	Peptoid
6	ZINC41398860	29	-74.74	Peptoid
7	ZINC04866612	30	-74.42	Peptide
8	ZINC40614900	27	-74.00	Peptide
9	ZINC28330572	31	-73.99	Peptoid
10	ZINC65683365	29	-73.78	Peptide
11	ZINC69022072	28	-73.71	Peptoid. Purchasable
12	ZINC41400402	29	-72.99	Peptoid
13	ZINC69022087	27	-72.79	Peptide-peptoid mix. Purchasable
14	ZINC36125115	27	-72.77	Peptoid. Purchasable
15	**ZINC71810038**	**28**	**-72.68**	**Peptoid. Purchasable**
16	ZINC41400074	29	-72.62	Peptoid
17	ZINC71810050	28	-72.46	Peptoid. Purchasable
18	ZINC62089977	28	-72.42	Peptoid
19	ZINC08346744	31	-72.10	Peptide-peptoid mix. Purchasable
20	ZINC35325177	29	-71.89	Peptide-peptoid mix. Purchasable
21	ZINC62187452	26	-71.63	Peptoid
22	ZINC22844848	28	-71.44	Peptide-peptoid mix. Purchasable
23	ZINC66885030	28	-71.42	Peptoid. Purchasable
24	ZINC24675306	29	-71.36	Peptide-peptoid mix. Purchasable
25	ZINC71962264	31	-71.33	Peptide. Purchasable
26	ZINC41400813	29	-71.15	Peptoid
27	ZINC71810037	28	-71.13	Peptoid. Purchasable
28	ZINC69022076	28	-70.98	Peptoid. Purchasable
29	ZINC41400547	28	-70.83	Peptoid
30	ZINC69021893	28	-70.80	Peptoid. Purchasable

The most promising compounds from Table 5 were then subjected to optimisation – see Table 6 and Figure 16 below, and Section 4.2.4 for further combinatorial optimisation with the use of peptoid library.

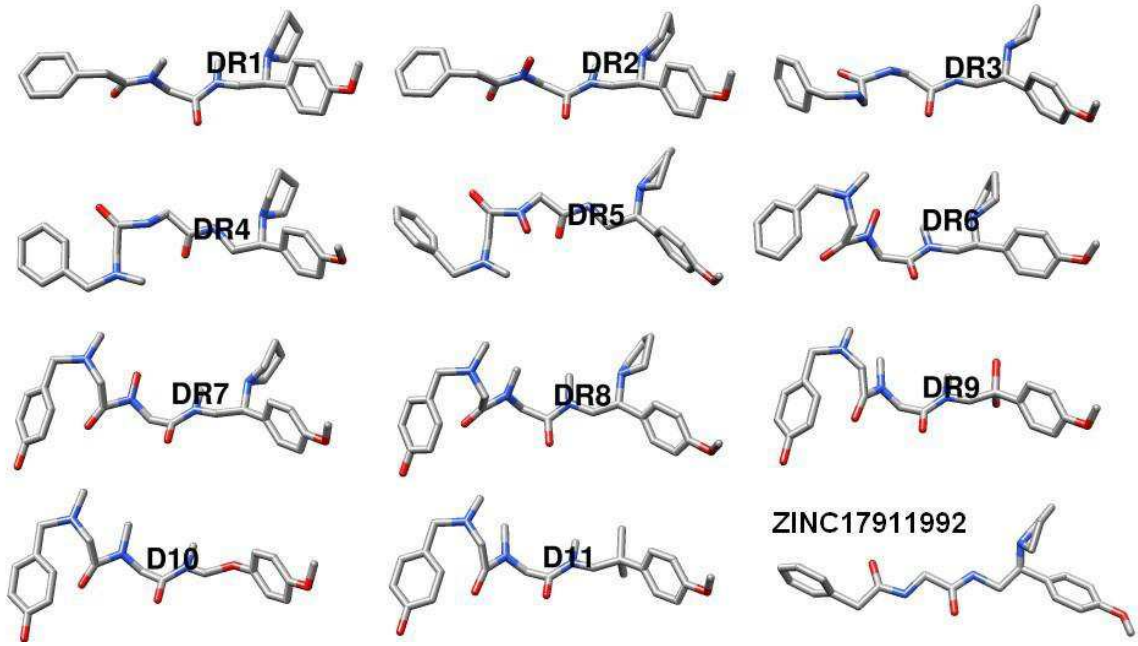

Figure 16 Lead optimisation: initial modifications of the ZINC17911992 compound

Table 6 Docking scores for modified derivatives of ZINC17911992 shown in Figure 16

Name	Score
DR1	-79.18
DR2	-81.02
DR3	-81.64
DR4	-81.26
DR5	-84.51
DR6	-84.39
DR7	-87.76
DR8	-90.41
DR9	-82.29
D10	-76.01
D11	-82.25
ZINC17911992	***-77.44***

Table 6 shows that small modifications of the ligand can lead to significant increase in binding energy. Further optimisation of this lead compound, and two other ones, was therefore performed with the use of the library of peptoids (see Section 4.2.4).

4.2.3 Drug repositioning

The by far most expensive and longest part of drug design process is the stage of the clinical trials. However, not all the drugs that have passed this phase are in use, either because of their limited therapeutic effect, or for other reasons such as marketing decisions of the pharmaceutical companies. It is also not unknown to use old drugs in new therapies. Examples include miltefosine, a drug initially intended for breast cancer and not proven successful in this respect, but now used for treating a parasitic disease visceral leishmaniasis; or thalidomide, the infamous hypnotic drug with teratogenic properties that found its new use as an immunomodulator in a variety of diseases [Aronson, 2007]. The compounds that have passed clinical tests have well-known toxicological and pharmacokinetic properties, safety profiles, and some are already approved for use in humans.

The ZINC database [Irwin *et al.*, 2012] contains a set of molecules approved by the US Food and Drug Agency (FDA), and this subset was also screened in search for thymidylate synthase inhibitors, both competitive (Table 4) and allosteric.

Losartan (Figure 17), an angiotensin II receptor antagonist drug used mainly in the treatment of high blood pressure, ranked well as a potential allosteric inhibitor of TS, docked to the binding site described by Cardinale *et al.* with docking score of -85.87 and large cluster of 18 conformations.

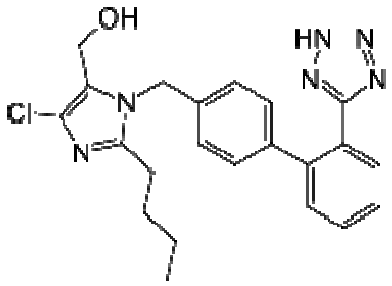

Figure 17 Chemical structure of losartan

It was therefore selected as a candidate for development. The molecule was parameterised according to the MD protocol described in Section 4.2.1 for hydroxyl analogues of dUMP, RESP charges were calculated and GAFF force field parameters assigned, and a 50-ns MD simulation of hTS-losartan complex in explicit TIP3P water was performed. Free energy of binding ΔG_{total} was estimated with MM-PBSA method (see Section 3.2.3) to be -52.4 ± 4.3 kcal/mol, which indicated that losartan may have inhibitory properties with respect to thymidylate synthase. More simulations are needed in order to estimate the behaviour of TS-losartan complex in the presence of the substrates.

In vitro experiments with the aim to estimate inhibition of human TS by losartan are currently being conducted by the team of W. Rode in the Nencki Institute of Experimental Biology.

4.2.4 Peptidomimetics

The use of polypeptides as drugs, despite many desirable attributes, is limited since they are *in vivo* rapidly degraded by proteases and may induce immune response. It is therefore necessary to find different chemical scaffolding that would function as peptide chains – peptidomimetics – and allow for the design of drugs with a structure similar to polypeptides, but resistant to proteases.

Peptoids – N-substituted glycine oligomers – are compounds with backbone similar to peptides, in which the side chain is shifted to the main chain nitrogen atom and the main chain carbon atom remains achiral (Figure 18). The monomers are linked by polyimide bonds and lack the amide hydrogen, thus preventing the formation of hydrogen bonds that in peptides stabilise the secondary structure.

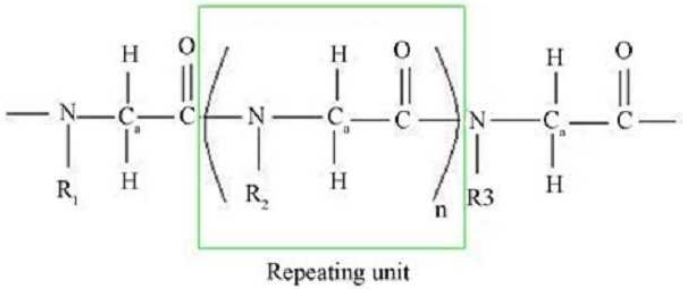

Figure 18 Primary structure of peptoid (illustration from [Nandel and Saini, 2010])

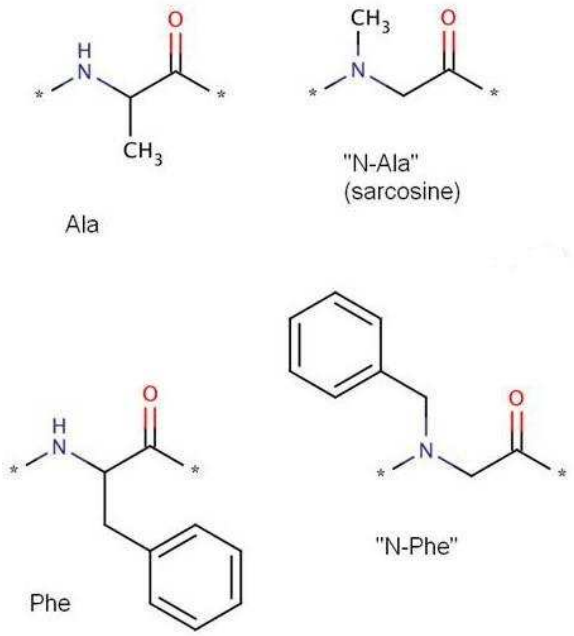

Figure 19 Examples of the structures of "N-amino acids" compared to amino acids

Unlike peptides (with the exception of peptide bonds to proline), peptoids may isomerise between *cis* and *trans* configuration of the amide bond (Figure 20), however, the interchange between configurations seems to be slower than receptor-ligand binding. Our initial computer simulations indicate that the cis-trans rotation rarely occurs in 300 K (Figure 21). In di- and tripeptoids studied experimentally by Sui *et al.* with the use of NMR spectroscopy, the population of *cis* conformation was about 35%.

Thus, a peptoid composed of *n* monomers may exist in solution as an equilibrium mixture of 2^{n-1} configurational isomers [Sui *et al.*, 2007].

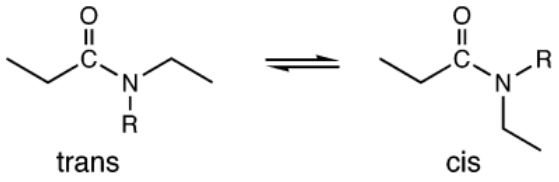

Figure 20 Cis and trans configuration of peptoid (illustration from [Sui *et al*., 2007])

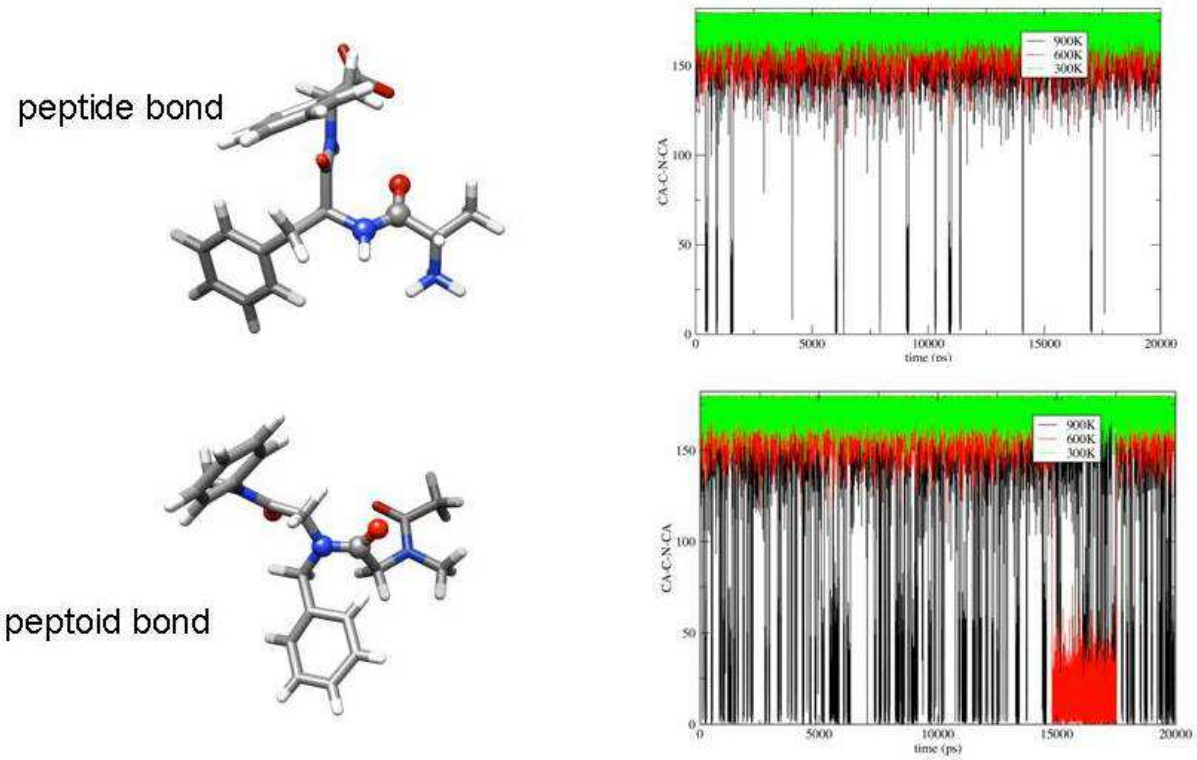

Figure 21 Molecular dynamics run (20 ns) in 900 (black), 600 (red) and 300 (green) K for the molecules with peptide bond and peptoid bond, Ca-C-N-Ca dihedral angle shown. Rotation around the peptoid bond occurs to a significant degree only in 900 K.

Computer simulations made by Nandel and Saini show that polypeptoid models adopt degenerate helical structures stabilised by carbonyl interactions. Apart from the use in drug design, these structures may also be useful in the design of surfactant molecules [Nandel and Saini, 2010].

Peptoids with various side chains are relatively easy to synthesise and that, together with the properties described above, makes them attractive targets for drug design. For example, it might be possible to design and synthesise peptoid molecules on the basis of peptides that inhibit thymidylate synthase *in vitro*, described in [Cardinale *et al.*, 2011].

4.2.4.1 Library of peptoids

The library of peptoids is a project carried out in cooperation with Dr. Agnieszka Bronowska (Heidelberg Institute for Theoretical Studies). The goal of this project was to create a library of N-substituted glycine units ("peptoids"), together with RESP

charges and PARM99 parameters, in the same way as amino acid units in Amber suite are constructed. These peptoid monomers can be connected to each other, which allows for easy "building" of putative ligands for docking and molecular dynamics studies. In this way, hypotheses on new peptidomimetic drugs can be easily tested *in silico*, leading to new, promising ligands that may be synthesised and tested experimentally.

Analogues of all proteinogenic amino acids present in eukaryotes were constructed, with the exception of glycine, which has no side chain, and proline, which is itself a secondary amine. Cysteine was included in two forms: regular protonated cysteine and cysteine involved in disulphide bridges and other bonds.

Nineteen structures of peptoid amino acid analogues were built with groups added to the ends: N-terminal $COCH_3$ group and C-terminal $NHCH_3$ group. The template structure is shown on Figure 22, where R denotes a side chain corresponding to the one present in the respective amino acid. All structures were in *trans* configuration. Additionally, calculations for *cis*-NAla and *cis*-NSer were performed in order to compare results and prove that the use of only one set of charges was justified.

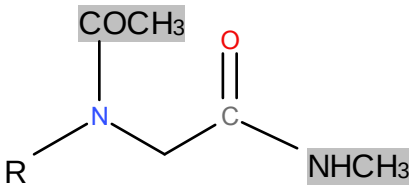

Figure 22 Initial structure of peptoid monomers used for geometry and RESP charges calculation

For all structures, geometry optimisation was performed with Gaussian09 suite [Gaussian09] at HF/6-31G* level, and charges fit to the electrostatic potential at the points selected according to the Merz-Singh-Kollman scheme were calculated. Then RESP charges were fitted with Antechamber programme (part of the Amber 11 suite [Amber11]), and Amber (PARM99 modified with frcmod.ff99SB) atom types were assigned so that the atom types corresponded to the ones known from the amino acids library.

Finally, the $COCH_3$ and $NHCH_3$ groups were truncated. If the unit lacked Amber angle or dihedral parameters due to an atom configuration not found in PARM99 data set, these parameters were assigned according to GAFF data set for atoms of the relevant type. The Amber library files for each unit were edited with respect to connectivity so that amide-N and carbonyl-C atoms, where the $COCH_3$ and $NHCH_3$ groups previously were (as shown in Figure 22), were assigned "head" and "tail" attributes, respectively.

Using these units, it is easy to build peptoid chains with Leap programme (part of the Amber 11 suite), ready for further simulations.

The following naming convention for the new residues was adopted. All the units are named NT*Z*, where N denotes peptoid unit, T – *trans* configuration (C would be *cis* configuration), and *Z* is an one-letter symbol of the corresponding amino acid.

Thus, charges and parameters were derived for the following residues: NTA, NTC, NTD, NTE, NTF, NTH, NTI, NTK, NTL, NTM, NTN, NTQ, NTR, NTS, NTT, NTV, NTW, NTX (disulphide bridge NCys), and NTY.

4.2.4.2 Further optimisation of the most promising compounds docked to the allosteric pocket of TS

From the compounds in Table 5, two were initially selected for further optimisation with the use of the library of peptoids: best scoring ZINC17911992 (already mentioned in Section 4.2.2), and ZINC71810038, suggested as a leading compound suitable for further modifications by a medicinal chemist (Figure 23).

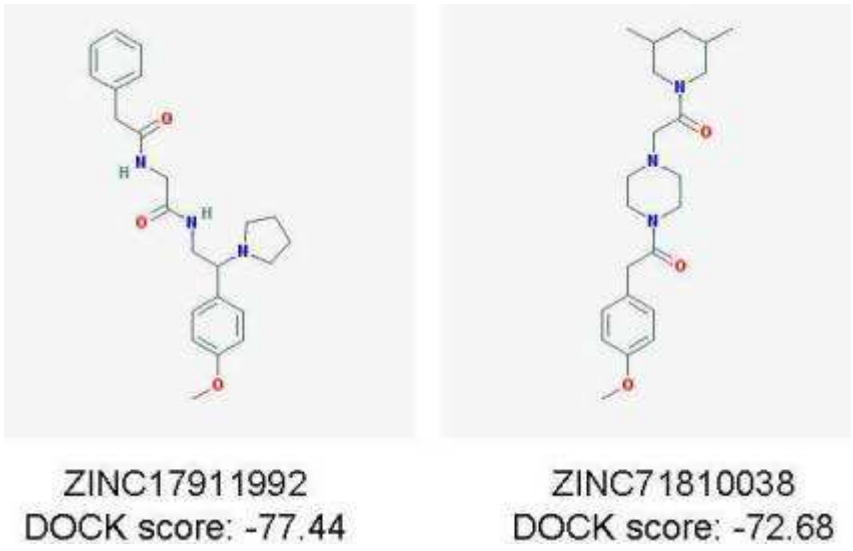

Figure 23 Compounds subject to further optimisation with the use of peptoid library

Derivatives of the ZINC71810038 molecule

On the basis of this compound, according to the suggestions of medicinal chemistry experts (informal discussions with Prof. M. Auer and his co-workers), eighteen molecules were constructed of the formula B2*Z*, where *Z* denoted a corresponding amino acid side chain, and B2 – a bulky molecular group, derivative of ZINC71810038 compound. An example with N-lysine side chain (B2K) is shown in Figure 24.

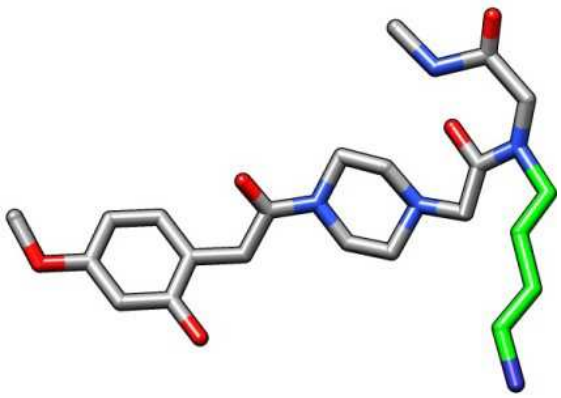

Figure 24 The structure of B2K molecule. The N-lysine side chain is marked green

The resulting docking scores are shown in Table 7. The suggested modifications improve the binding properties of the compound.

Table 7 Potential leads (allosteric) based on ZINC71810038 modified with amino acid side chains

Name	DOCK Score
B2A	-73.30
B2C	-75.78
B2D	-79.92
B2E	-79.62
B2F	-78.02
B2H	-73.41
B2I	-77.36
B2K	-88.91
B2L	-80.38
B2M	-80.43
B2N	-78.34
B2Q	-76.23
B2R	-80.06
B2S	-74.86
B2T	-76.81
B2V	-77.25
B2W	-84.44
B2Y	-82.62

A combinatorial set of peptidomimetic ligands for thymidylate synthase: derivatives of ZINC17911992 molecule

As the results of the modifications of ZINC17911992 molecule were promising (Table 6), this approach was taken further and a series of over 300 combinatorially modified structures with short, two-peptoid chains was designed (Figure 25). An example of the series is shown in Figure 26.

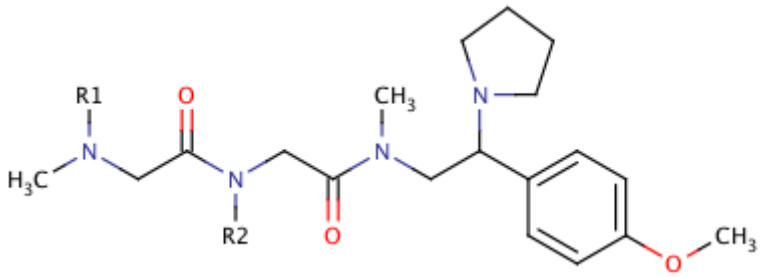

Figure 25 Scheme of the set. R1 and R2 denote amino acid side chains

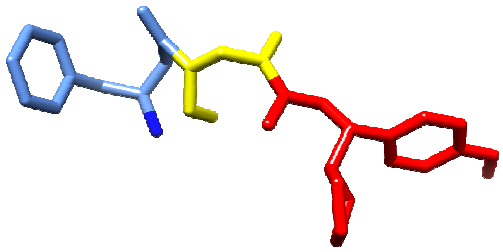

Figure 26 Example of P–R1–R2–B series of ligands: PFSB; CH_3(dark blue)–NPhe(blue)–NSer(yellow)–B(red). B denotes the bulky "tail" from the ZINC17911992 molecule

The series of 324 ligands was then docked to the allosteric pocket of thymidylate synthase, and the best 30 results are presented in Table 8.

Table 8 Potential leads (allosteric) among combinatorially modified ligand series based on ZINC17911992. Best results are highlighted.

Sequence	Score	Sequence	Score
PKTB	-102.18	**PKFB**	**-109.63**
PKVB	**-109.21**	**PHKB**	**-133.88**
PKSB	-97.55	**PKKB**	**-127.85**
PKLB	**-111.44**	PKDB	-99.50
PKAB	**-110.85**	PEKB	-97.93
PKDB	**-111.51**	PFKB	-97.14
PKRB	-100.65	PIKB	-97.91
PKCB	-100.61	PKQB	-96.89
PKNB	**-109.13**	PFRB	-96.66
PKMB	**-104.04**	PERB	-95.53
PKHB	-100.31	PHRB	-95.63
PKYB	**-107.28**	PDKB	-95.98
PKWB	-99.62	PLKB	-98.42
PKEB	**-103.62**	PLRB	-95.02
PKIB	-95.68	PMKB	-96.79

The reference score for these compounds is the LSCQLYQR peptide (calculated as -107.93), docked to the same allosteric pocket, which is a TS inhibitor *in vitro* [Cardinale *et al.*, 2012].

The best scoring ligands will be further investigated in molecular dynamics simulations with the use of the library of peptoids we created.

5 Other work

The author's scientific interests involved also quantum chemical calculations with the aim to better understand and interpret experimental results. Several papers have been published in this regard, presented here in approximately chronological order and attached to the dissertation.

5.1 DFT studies of interactions between nucleic acid bases

To understand molecular aspects of genetic processes, it is important to gain data on the interaction energy between nucleic acids in different mutual arrangements, and on the barriers of transition from one arrangement to another. Experiments usually only yield information on equilibrium states; it is therefore necessary to use computational methods to predict geometries and interaction energies for the local minima on the potential energy surface of base pairs. In the work carried out with V.I Poltev, V.I Danilov, A. Leś *et al.* local minima for DNA fragments of various size – from base pairs to deoxydinucleoside monophosphate anions – were found with molecular mechanics calculations and then investigated thoroughly by DFT studies. It was concluded that the parallel arrangement of the bases in B-like single-stranded nucleic acid conformations is dictated by the sugar-phosphate backbone rather than by the "stacking" interaction between base rings. Moreover, the results indicate that the nearly parallel arrangement of the base rings is already present in the single DNA strand [Poltev *et al.*, 2008; Poltev *et al.*, 2009] (Full text in Appendices 2 and 3).

5.2 Boron analogues of nucleosides

Computational analysis of the regio- and chemoselectivity of N-vinylurea and its derivatives in the synthesis of novel 5,6-dihydroborauracil derivatives

Boron analogues of biologically active compounds have interesting properties, supposedly associated with the electron deficiency of sp^2 boron moiety that may bind to an enzyme, which would lead to the formation of a stable sp^3 „boronate" complex and the inhibition of the enzyme. In addition to this, boron-containing compounds can be useful in boron-neutron capture therapy (BCNT). Borauracils, borathymines or boracytosines, and their derivatives, could play an important role in the search for new biologically active compounds, e.g. anticancer drugs. However, to date only a few boron analogues of nucleic acid bases are known.

In the work of T. Ruman *et al.* (2009) the synthesis of novel boron compounds – 5,6-saturated borauracil derivatives (4-bromo-5,6-dihydroborauracil, 4-hydroxy-5,6-dihydroborauracil and 4-methoxy-5,6-dihydroborauracil) is presented along with the synthesis of other boron compounds from N-vinylurea.

The computational part of this work concerned the selectivity of the substrates, N-vinylurea and its derivatives, in hydroboration reaction. The calculated natural

charges indicate that, in agreement with the experimental results, the partially negative hydrogen atom of the borane complex should attack the carbonyl carbon of N-vinylurea, which confirms the proposed mechanism of the reaction. The theoretical (DFT) calculations were performed in *Gaussian03* suite at the B3LYP/aug-cc-pVTZ level. The electronic density distribution has been analysed with the use of the NBO 5.G module also implemented in *Gaussian03* [Ruman *et al.*, 2009a] (Full text in Appendix 4).

Calculation of ^{1}H NMR parameters of the hydroxyborohydride anion

The team of T. Ruman *et al.* (Rzeszow University of Technology) described a fast, economic method of the synthesis of sodium and potassium salts of hydroxyborohydride $(BH_3OH)^-$ anion. The ^{1}H NMR spectrum of $(BH_3OH)^-$ was recorded for the first time and NMR studies showed higher reactivity of hydroxyborohydride towards acetone carbonyl than borohydride.

The calculations conducted by the author of the dissertation and Prof. A. Leś involved computations of spin-spin coupling constants ^{1}J(B–H) for $(BH_3OH)^-$, $(BH_3OH)Na$, BH_4^-, BH_3, and B_2H_6 at the B3LYP/aug-cc-pVTZ level. Good agreement with experimental data confirmed the hypothesis that the $(BH_3OH)^-$ anion was detected as an individual species in the solution. The new, high yield synthetic route of hydroxyborohydride salts makes them attractive as reagents [Ruman *et al.*, 2007] (Full text in Appendix 5).

5.3 Modification of stavudine

DFT calculations of the geometric and electronic properties of the tetrahedral boric acid complexes – derivatives of stavudine

Stavudine (2',3'-didehydro-3'-deoxythymidine, Figure 27) is a synthetic analogue of thymidine approved for use in the European Union and USA for treatment of HIV infections in combination with other antiretroviral compounds.

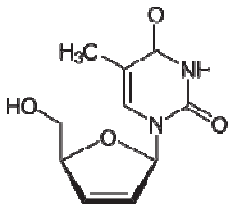

Figure 27 Stavudine

The active form of the drug, stavudine 5'-triphosphate, is a competitive vs. dTTP inhibitor of HIV nucleoside reverse transcriptase that causes the termination of DNA polymerisation. In search of new antiretroviral compounds, two boron-containing analogues of stavudine 5'-phosphate were synthesised (Figure 28). In these compounds, in place of the phosphate group, the boron moiety forms complexes in which an additional electron pair donor coordinates to boron atom, and the resulting slightly distorted tetrahedral group mimics the geometric and electronic properties of the phosphate group. This could possibly improve the cell wall transfer properties and increase the bioavailability of stavudine.

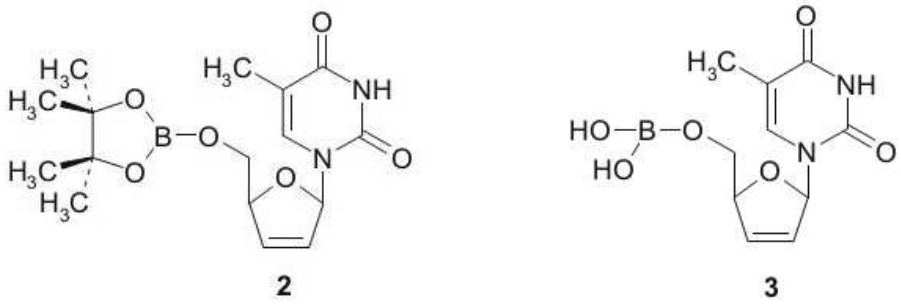

Figure 28 Boron-containing derivatives of stavudine

Compound **2** was synthesised by direct reaction of stavudine (also synthesised during this study) with large excess of pinacolborane (4,4,5,5-tetramethyl-[1,3,2]dioxaborolane), and compound **3** was formed in the reaction of stavudine with borane-dimethylsulphide complex, followed by hydrolysis of the resulting hydride compound 5'–CH_2–O–BH_2. The ^{11}B spectra of **2** and **3**, recorded in non-coordinating solvent, show broad resonances at 22.4 and 17.8 ppm, respectively, in accordance to the previously known data. Addition of H_2O into the NMR tube containing the compound **2** resulted in an *upfield* shift of ^{11}B resonance up to –11.2 ppm. This is caused by the formation of a stable eight-electron, tetrahedral adduct as a result of the coordination of an electron pair donor (such as water, thiol, amine, alcohol, hydroxyl anion, or carboxylate anion) to the positively polarised boron atom. Similar behaviour was also observed when H_2O was added to **3**. No significant changes were observed in the ^{1}H NMR spectrum during H_2O titration, which indicates the lack of side reactions.

These tetrahedral boron moieties were considered to be potentially capable of mimicking the steric and electronic properties of the phosphate group. In order to test this hypothesis, DFT calculations were performed at B3LYP/LANL2DZ level with the use of *Gaussian 03* suite. In the course of these calculations, three models of 5'-CH_2-O-X, representing stavudine and two adducts of the compound **3**, were taken into account (Figure 29).

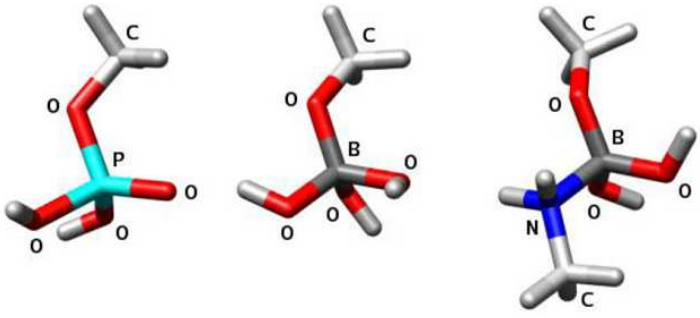

Figure 29 The models of stavudine-5'-CH_2-O-X (nucleoside group is truncated and replaced with methyl group); X=phosphate (left), –$B(OH)_3$ (middle), and –$B(OH)_2$:NH_2CH_3 (right); geometries after optimisation

The geometries were optimised and Merz-Kollman ESP charges and molecular volume calculated. All models adopt tetrahedral geometry with similar bond lengths (slightly shorter in the case of borane moieties) and angles. The values of atomic charges and

molecular volumes are also close. The comparison between phosphate and borane moieties suggests that their local properties should lead to similar short- and long-range interactions with the environment [Ruman *et al.*, 2010a] (Full text in Appendix 6).

5.4 Non-enzymatic modifications of thymidylate synthase

Thiophosphoramidates as agents for studies on phosphorylation/dephosphorylation of proteins

The use of thiophosphate compounds offers insight into the processes of phosphorylation of amino acids. Compared to phosphates, thiophosphates have an additional centre of chirality on phosphorus, are more stable towards chemical and enzymatic hydrolysis, and are easier to detect by analytical methods. The preparation of thiophosphoramidate-modified biomolecules under mild conditions that preserve protein structure could be useful for studying the biological function of the corresponding phosphoramidates.

The team of T. Ruman *et al.* performed the synthesis of ^{15}N-thiophosphoramidates (potassium and diammonium salts), and tested their ^{15}N and ^{31}P NMR properties as well as potential to thiophosphorylate histidine, thymidine, glucose and 2-deoxyribose.

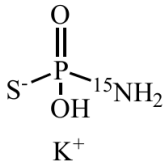

Figure 30 ^{15}N-enriched potassium thiophosphoramidate

A series of DFT calculations was conducted in order to estimate the ^{31}P NMR chemical shifts of thiophosphoramidates by analysing electronic shielding of thiophosphoramidate and thiophosphate ions. The calculations were performed using Gaussian03 suite at the B3LYP/aug-cc-pVTZ level. The aim was to estimate trends rather than reproduce exact values of NMR parameters, which in the case of ^{31}P NMR remains a long-term challenge. It was concluded that DFT methods can be useful in anticipating the NMR chemical shifts of newly synthesised thiophosphate compounds [Ruman et al., 2009b] (Full text in Appendix 7).

Thiophosphorylation of free amino acids and enzyme thymidylate synthase

Potassium and diammonium thiophosphoramidates in Tris- and ammonium carbonate based buffer solutions were used to thiophosphorylate free amino acids as well as recombinant human and *C. elegans* thymidylate synthase. This reaction environment proved to be non-destructive for protein function – the thiophosphorylation reaction lowered the enzyme activity by ~50% after 24 hours.

A series of phosphorylated and thiophosphorylated amino acid derivatives was prepared, their ^{31}P NMR spectra were recorded, and computational (DFT) estimations of chemical shifts were made in order to properly assign NMR resonances of the thiophosphorylated protein to particular thiophosphorylated amino acid side chains. The presence of thiophosphorylated Lys, Ser and His moieties was proven in this way.

Moreover, the ^{31}P resonances of thiophosphorylated amino acids are shifted *downfield* by roughly 42 ppm on average, compared to the corresponding phosphate compounds, which allows for a rough estimate of chemical shift for a thiophosphate when this parameter for a corresponding phosphate is known.

The DFT calculations (geometry optimisation and calculation of ^{31}P NMR shielding constants) were again performed at the B3LYP/aug-cc-pVTZ level with *Gaussian03* suite. To save computational time, the structures of thiophosphorylated (TP) and phosphorylated (P) amino acids were truncated in the regions remote from the (thio)phosphate group (Figure 31).

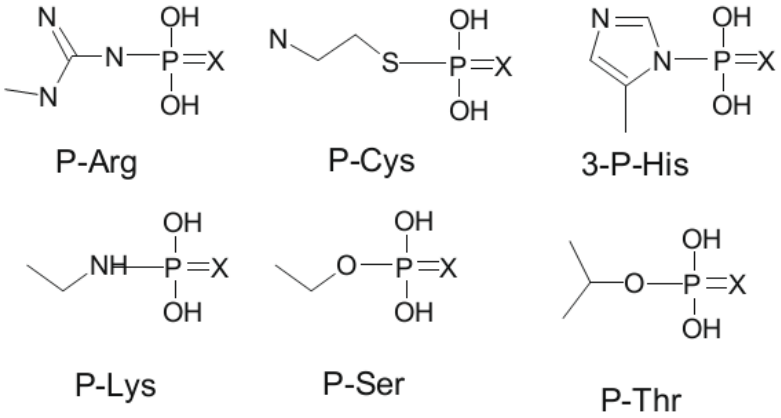

Figure 31 Models of phosphorylated (X=O) and thiophosphorylated (X=S) amino acids

Due to the complexity of the interactions between the analysed compounds and buffer ingredients, and to the unreliability of concentration estimation of neutral, mono- and dianionic forms of phosphate in the solution, it is difficult to properly set a reference value for the calculation of ^{31}P NMR chemical shift $\delta_P = \sigma_{ref} - \sigma_x$ (σ being the isotropic shielding constant) with phosphoric acid as a reference compound. Moreover, as mentioned before, the calculations of shielding constants with the method used give a trend assessment rather than exact values. Thus, chemical shifts were first estimated as differences between the calculated isotropic shielding constants for neutral forms of P- or TP-amino acids and non-dissociated H_3PO_4. Then, after comparison with experimental results available for TP-His, TP-Lys, and TP-Ser, an additional factor of 33.7 ppm was subtracted from all calculated chemical shifts. This way, values of ^{31}P chemical shifts were suggested for TP-Arg and TP-Thr, for which experimental data were not available. In addition to this, the results for TP-Cys indicated the presence of both “thione” (P=S) and “thiol” (P-SH) tautomeric forms.

The experimental and computational results outlined above enabled the identification of thiophosphorylated lysine, serine and histidine moieties in the thiophosphorylated recombinant thymidylate synthase. Methods presented in the paper can be useful in predicting ^{31}P NMR chemical shifts of thiophosphorylated compounds [Ruman *et al.*, 2010b] (Full text in Appendix 8).

Studies on tyrosine nitration of thymidylate synthase

Nitration of protein tyrosine appears endogenously in diseases involving increased biosynthesis of NO, including cancer. This reaction may lead to the loss or gain of

activity of the protein. The effect of thymidylate synthase nitration was studied with respect to kinetic and physicochemical (NMR and MS) properties, on human, mouse and *C. elegans* recombinant TSes, nitrated *in vitro*. In order to interpret the obtained NMR resonances, model tripeptides of a general formula H_2N-Gly-X-Gly-COOH were nitrated and analysed by NMR spectroscopy, with X=Phe, Tyr, Trp, Lys, Arg, His, Ser, Thr, Cys, and Gly. Theoretical calculations of ^{15}N NMR chemical shifts for the models of the nitrated amino acids were also performed in the same way as described above for the calculations of ^{31}P NMR parameters, using truncated models (see Fig. 3 and Table 1 in Appendix 9), and the results were compared to experimental spectra.

The nitration of TS did not seem to influence binding to dUMP, FdUMP or mTHF, as reflected by K_m values, however, the enzyme catalytic potency was affected, with 2-fold decrease of V_{max}^{app}.

The comparison of NMR parameters of nitrated thymidylate synthase with the NMR analysis of model tripeptides, as well as the results of MS experiments, allowed for the identification of nitrated tyrosine. The presence of other nitrated moieties could not be confirmed [Dąbrowska-Maś et al., 2012] (Full text in Appendix 9).

6 Discussion, conclusions and future outlook

Rationalisation of observed binding properties of a set of dUMP analogues to thymidylate synthase

Three selected analogues of dUMP, substituted at C5 position with small functional groups: –OH, $–CH_2OH$, and $–OCH_3$, were investigated computationally. The dynamical behaviour of the ternary complexes comprising of thymidylate synthase monomeric unit, the cofactor methylenetetrahydrofolate in its active, open form, and the ligand, was analysed in the course of 20 ns molecular dynamics runs. For comparison and verification of the results, analogical calculations with dUMP and a known TS inhibitor, FdUMP, were also performed. The estimation of binding energies (enthalpic contribution) made with the use of MM-PBSA method (Table 1) agreed qualitatively with the experimental results given according to [Jarmuła *et al.*, 2000]. Apart from the fluorouracil derivative, the most strongly bound and most ordered ligand in the active site is dUMP-OH. Figure 14 shows the impact of hydrogen bonds network on the ligand/protein binding capabilities. The more bulky and flexible substituent $–CH_2OH$ remains less ordered and the binding is weaker, as is in the case of $–OCH_3$ substituent which has no additional hydrogen bonds. The obtained results of the molecular simulations strongly indicate that these factors – namely, the size of the substituent and its hydrogen bonding ability – affect heavily the binding properties of these ligands. Qualitative agreement between experimental and computational data gives confidence in the conclusions made of the simulation results. Therefore, these should be taken into account during the design of new, C5-substituted uracil analogues intended as inhibitors of thymidylate synthase or agents for enzyme-catalysed therapies.

Allosteric inhibition of TS as an alternative to overcome drug resistance

In addition to causing some adverse effects, the known orthosteric, i.e. active site inhibitors of thymidylate synthase, may induce drug resistance. Human thymidylate synthase is believed to be self-regulated by binding with its own mRNA when the ligands are not bound. It is postulated that due to conformational change upon ligand binding the TS loses this self-regulation capacity, which in turn leads to the overexpression of the protein (see Section 1.2.5). Moreover, the efficiency of TS degradation decreases when the ligands remain bound. As an alternative to competitive inhibitors, allosteric thymidylate synthase inhibitors that bind to another site, e.g. at the dimer interface, are investigated. These compounds preferably would prevent the ligands from entering the active site of the protein without disrupting self-regulation. A putative allosteric binding site at the TS dimer interface was found by Cardinale *et al*. (2011). The available X-ray structures showing ligands binding to this site include a TS-peptide complex described in [Cardinale *et al.*, 2011], PDB access code 3N5E, and a TS-purpurogallin complex (Celeste and Lebioda, unpublished work, PDB access code 4GD7).

In the course of this work a plethora of ligands binding to the above described site was investigated computationally in order to select promising candidates for further optimisation and experimental studies. The compounds under study included existing drugs, prodrugs and naturally occurring compounds from databases such as ZINC and Afro-DB, and newly designed compounds (peptidomimetics).

Peptoids as allosteric TS inhibitors

Peptoids (see Section 4.2.4) were brought into drug discovery projects in the early 1990s and they remain a developing class of peptidomimetics since then. While peptides and their binding to therapeutically-relevant protein targets are well-studied in the framework of medicinal chemistry and drug design, the development of peptide-based therapeutics is hindered by their poor absorption, *in vivo* degradation due to proteolytic digestion, rapid clearance, and their potential for aggregation, misfolding, and immunogenicity. Unlike peptides, peptoids are resistant to denaturation due to their lack of backbone H-bonds, and are very straightforward to synthesise, with readily available (and cheap) precursors and lack of chirality in the backbone. Yet their side chains may – if required – resemble very closely those of peptides and proteins. Thus, peptoids have demonstrated usefulness for medicinal chemistry as protein mimics and as potential replacements for small-molecule drugs.

In the recent years, several applications of short peptoids as potent antimicrobial and anticancer agents have been reported. For instance, Chongsiriwatana et al. (2011) published the synthesis and evaluation of peptoid analogue of antimicrobial lipopeptides with alkyl chains, consisting of only 5 peptoid monomers, with selective, broad-spectrum antimicrobial activity, and as potent as the antimicrobial peptide pexiganan, which is a 22-residue long analogue of magainin-2 [Chongsiriwatana et al. (2011)]. Earlier, Mas-Moruno et al. (2007) reported the design of libraries of small peptoids with antiproliferative activity against representative human neoplastic cell lines. These were

synthesized and successfully assayed for their biological activity [Mas-Moruno *et al.*, 2007]. These peptoids were found cytotoxic and were reported to be the smallest peptidomimetics with such activity. Due to their properties, these compounds did not suffer from diketopiperazines (DKP) formation, a problematic side reaction typically observed in peptide synthesis. Moreover, these peptoids were conveniently synthesized on a solid surface (BAL resin).

Therefore, the idea of using peptoids as allosteric inhibitors of TS is very appealing.

The screening of compounds from ZINC database also revealed several peptoid compounds that can be a basis for optimisation and design of novel drugs (see Table 5 and Section 4.2.4).

Virtual screening of FDA-approved drugs and peptoid ligands present in ZINC database

Molecular docking studies conducted in order to find new candidates for TS inhibitors, both orthosteric and allosteric, involved over eight thousand compounds (see Section 4.2.2), among which were around fifteen hundred FDA-approved drugs and nutriceuticals, and a similar number of natural products from AfroDB and NuBBE databases. The docking procedure was calibrated with the use of known TS inhibitors and non-binding compounds, and in the case of compounds targeting the allosteric site, the reference structure was the complex of human TS bound to a short peptide, described by Cardinale *et al.* (2011).

This extensive screening resulted in many hits (see Tables in Section 4.2.2), some of which were further optimised (results shown e.g. in Table 6 and Section 4.2.4, Table 8) and yielded promising leads for more thorough molecular dynamics studies and experiment.

Drug repositioning is also an important part of the drug design process. A promising hit from FDA-approved set, losartan – a drug used in the treatment high blood pressure – was investigated more thoroughly as a putative allosteric inhibitor of thymidylate synthase, by molecular dynamics simulations (see Section 4.2.3).

Peptoids as building blocks for rational drug design

While molecular simulations of peptides and proteins are standard and all molecular simulation packages have implemented tools to perform those, relatively few tools suitable for simulations of peptoids and peptoid-protein systems are available.

A set of model peptoid structures, studied by Butterfoss *et al*. using QM methods, as well as three experimentally determined peptoids, were simulated using Amber force fields [Butterfoss *et al.*, 2009, Voelz *et al*., 2011]. Their results showed that Amber force fields, when used with a GBSA implicit solvation model, successfully reproduced the peptoid torsional landscape as well as the major conformers of known peptoid structures.

Very recently, Mirijanian *et al.* published an atomistic force field for peptoids implemented in the CHARMM22 package [Mirijanian *et al.*, 2014]. In their work, the key parameters for peptoids were adjusted to match both experimental data and

quantum mechanical calculations for two model peptoid systems – namely, disarcosine and dimethylacetamide.

Despite these encouraging reports, none of the published works to date reported on the development of a set of standard, residue-like building peptoid blocks, suitable for setting up the molecular simulations, in a manner similar to peptides, proteins, and nucleic acids.

The results presented in this dissertation cover this need and herein the collection of ready-to-use peptoid residues, with charges fitted and atom parameters assigned, is reported. These additional parameters have been developed for Amber package (AMBER11 version and beyond) and can be used standalone as well as in peptoid-peptide and peptoid-protein systems.

The test MD simulations performed in explicit (TIP3P) as well as implicit (GB) solvents correctly predicted the structure of cyclic peptoid octamers: 8-sarcosine and cyclosporine A (in a complex with cyclophilin B) to an accuracy of 1.2 Å backbone RMSD.

Overall, these results show that currently available molecular modelling techniques can successfully perform *de novo* structure prediction of short peptoids and reliably study the dynamics and binding of peptoid ligands to their cognate protein receptors. Moreover, the results reported herein provide a potentially valuable tool for the rational design of drug-like peptoids with desired structural and dynamic properties.

"Combinatorial" improvement of the leads selected from ZINC database

Around 300 compounds with peptoid bonds found in the ZINC database were docked to the dimer interface to select potential allosteric inhibitors of thymidylate synthase. Several leads for synthesis and further development were selected (Table 5 in Section 4.2.2). Two selected compounds were then optimised combinatorially by adding amino acid side chains to peptoid units (Section 4.2.4.2). This optimisation procedure led to significant improvement of docking scores for several derivatives of the original compounds (Table 8). In order to assess computationally the actual inhibitory potential of the selected derivatives, molecular dynamics simulations are required. Setting up of these simulations will be easy with the use of the library of peptoids that was developed in the course of this work (Section 4.2.4.1). This set of parameters for Amber (and other molecular dynamics packages where Amber force fields are used) comprises of 19 peptoid analogues of amino acids that can be used as "building blocks" for quick construction and parameterisation of peptidomimetic structures for the purposes of molecular dynamics simulations, in the same way as peptide chains are constructed in Amber.

Molecular dynamics simulations set up with the use of the above mentioned library of peptoids will be performed for complexes of TS with several peptoid ligands selected through combinatorial improvement and molecular docking procedure described in Section 4.2.4.

Quantum mechanical studies

The development of modern computational tools – e.g. numerical methods such as linear scaling algorithms, or semiempirical methods – made feasible the studies of even large systems at the quantum mechanics level. The *ab initio* QM methods also have numerous applications in the field of studying biological molecules. In this work they were essential for setting up molecular dynamics simulations (determination of atomic charges, see Sections 3.2.2 and 4.2.1), and in studying reaction mechanisms in small models as well as in a QM/MM setting, where the active centre of the enzyme is subjected to QM calculations and the surrounding protein is studied with semi-empirical or MM methods (Section 4.1). Moreover, QM calculations are useful in predicting NMR parameters of newly synthesised compounds and therefore in the identification of the product of a reaction such as thiophosphorylation or nitration of the protein (Section 5.4). In this case, the use of a relatively quick DFT method (B3LYP/aug-cc-pVTZ) applied to the truncated models of molecules under study allowed for a qualitative assessment of NMR parameters of the products of a reaction and contributed to the identification of these products.

Most experimental methods, with the exception of X-ray analysis, do not yield information on the geometry of the reagents and on interaction energies of the individual components of the system. Quantum mechanical calculations, in particular DFT methods, may fill these gaps in the understanding of processes in biological systems. In the course of this work the geometry as well as the potential short-range interactions (atomic charges, molecular volume) of novel boron-containing derivatives of stavudine was studied computationally (Section 5.3), complementing experimental results.

QM methods are also essential in the theoretical studies on nucleic acids. They enable an assessment of the interactions between nucleic acid bases in DNA, giving information on local minima of geometry, mutual arrangements of the bases, and on the stability of the DNA strands (Section 5.1).

Conclusion

Computational methods of chemistry, from *ab initio* QM calculations to large-scale molecular dynamics simulations and high-throughput molecular docking protocols, found their use in studying biochemical processes and rational drug design. In this dissertation an overview of some of these uses is presented in relation to the enzyme thymidylate synthase and the search of therapies for TS-related cancers. Molecular docking studies can help in finding drug candidates for further development; the library of peptoids is intended to facilitate the development of novel peptidomimetic compounds with therapeutic potential; molecular dynamics simulations are needed to investigate the studied compounds thoroughly and establish the dynamics, binding properties and mutual arrangements of the components of the system; and *ab initio* quantum mechanical calculations allow for investigating the processes of interest at the atomic level and contribute to the interpretation of the experimental data.

7 Acknowledgements

The computational work described here was performed on the supercomputers and computer clusters in:

- Interdisciplinary Centre of Modelling, University of Warsaw
- Heidelberg Institute for Theoretical Studies, Heidelberg, Germany
- Faculty of Chemistry, University of Warsaw.

8 References

[Amber8] D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, B. Wang, D.A. Pearlman, M. Crowley, S. Brozell, V. Tsui, H. Gohlke, J. Mongan, V. Hornak, G. Cui, P. Beroza, C. Schafmeister, J.W. Caldwell, W.S. Ross, and P.A. Kollman (2004), AMBER 8, University of California, San Francisco.

[Amber11] D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman (2010), AMBER 11, University of California, San Francisco.

[Aronson, 2007] Aronson JK. Old drugs – new uses. *Br J Clin Pharmacol*. 2007 Nov;64(5):563-5.

[Bailey and Ayling, 2009] Bailey SW, Ayling JE, "The extremely slow and variable activity of dihydrofolate reductase in human liver and its implications for high folic acid intake", *Proc Natl Acad Sci U S A* 106(36): 15424–15429, 2009

[Bayly *et al.*, 1993] Bayly CI, Cieplak P, Cornell WD, Kollman PA. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.*, 1993, 97 (40), pp 10269–10280

[Butterfoss *et al.*, 2009] Butterfoss GL, Renfrew PD, Kuhlman B, Kirshenbaum K, Bonneau R. A preliminary survey of the peptoid folding landscape. *J Am Chem Soc*. 2009 Nov 25;131(46):16798-807.

[Bernstein, 2000] Bernstein HJ. Recent changes to RasMol, recombining the variants. *Trends Biochem Sci*. 2000 Sep;25(9):453-5.

[Cardinale *et al.*, 2011] Cardinale D, Guaitoli G, Tondi D, Luciani R, Henrich S, Salo-Ahen OM, Ferrari S, Marverti G, Guerrieri D, Ligabue A, Frassineti C, Pozzi C, Mangani S, Fessas D, Guerrini R, Ponterini G, Wade RC, Costi MP, "Protein-protein interface-binding peptides inhibit the cancer therapy target human thymidylate synthase", *Proc Natl Acad Sci U S A* 108(34): E542-9, 2011

[Carreras and Santi, 1995] Carreras CW, Santi DV. The catalytic mechanism and structure of thymidylate synthase. *Annu Rev Biochem.* 1995;64:721-62.

[Cheng *et al.*, 2009] Cheng T, Li X, Li Y, Liu Z, Wang R. Comparative assessment of scoring functions on a diverse test set. *J Chem Inf Model.* 2009 Apr;49(4):1079-93.

[Chongsiriwatana *et al.*, 2011] Chongsiriwatana NP, Miller TM, Wetzler M, Vakulenko S, Karlsson AJ, Palecek SP, Mobashery S, Barron AE. Short alkylated peptoid mimics of antimicrobial lipopeptides. *Antimicrob Agents Chemother.* 2011 Jan;55(1):417-20.

[Cornell *et al.*, 1993] Cornell WD, Cieplak P, Bayli CI, Kollman PA. Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.*, 1993, 115 (21), pp 9620–9631

[Dąbrowska-Maś *et al.*, 2012] Dąbrowska-Maś E, Frączyk T, Ruman T, Radziszewska K, Wilk P, Cieśla J, Zieliński Z, Jurkiewicz A, Gołos B, Wińska P, Wałajtys-Rode E, Leś A, Nizioł J, Jarmuła A, Stefanowicz P, Szewczuk Z, Rode W. Tyrosine nitration affects thymidylate synthase properties. *Org Biomol Chem.* 2012 Jan 14;10(2):323-31.

[DrugBank] http://beta.drugbank.ca, Open Data Drug & Drug Target Database

[Dzik *et al.*, 1997] Dzik JM, Zieliński Z, Jarmuła A, Rode W, "Interactions of Thymidylate Synthase with 5-Hydroxy-dUMP and 5-Hydroxymethyl-dUMP, and Their 4-Thio Analogues", *Chemistry and Biology of Pteridines and Folates*: 415-418, 1997

[Finer-Moore *et al.*, 2003] Finer-Moore JS, Santi DV, Stroud RM, „Lessons and Conclusions from Dissecting the Mechanism of a Bi-substrate Enzyme: Thymidylate Synthase Mutagenesis, Function and Structure", *Biochemistry* 42: 248-256, 2003

[Fusaro *et al.*, 2010] Fusaro M, Jurkiewicz A, Jarmuła A, Leś A, Rode W, "Hypothesis of a proton switch in QM/MM modelling of interaction of dUMP analogues with thymidylate synthase", *Molecular Simulation* Volume 36, Issue 13, **2010**

[Garg *et al.*, 2010] Garg D, Henrich S, Salo-Ahen OM, Myllykallio H, Costi MP, Wade RC, "Novel approaches for targeting thymidylate synthase to overcome the resistance and toxicity of anticancer drugs", *J Med Chem.* 53(18):6539-49, 2010

[Gaussian03] Gaussian 03, Revision C.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Wallingford CT, 2004**.**

[Gaussian09] Gaussian 09, Revision A.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2009.

[G09 User's Reference] Gaussian 09 User's Reference, available online at http://www.gaussian.com/g_tech/g09ur.htm, description of the Opt section and Berny optimisation algorithm can be found at http://www.gaussian.com/g_tech/g_ur/k_opt.htm

[Humphrey *et al.*, 1996] Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph*. 1996 Feb;14(1):33-8, 27-8.

[Hyatt *et al.*, 1997] Hyatt DC, Maley F, Montfort WR, "Use of strain in a stereospecific catalytic mechanism: crystal structures of Escherichia coli thymidylate synthase bound to FdUMP and methylenetetrahydrofolate", *Biochemistry* 36(15):4585-94, 1997

[Irwin *et al.*, 2012] Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG. ZINC: a free tool to discover chemistry for biology. *J Chem Inf Model*. 2012 Jul 23;52(7):1757-68.

[Jakalian *et al.*, 2002] Jakalian A, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem*. 2002 Dec;23(16):1623-41.

[Jarmuła *et al.*, 2000] Jarmuła A, Leś A, RodeW, "Different Activities of 5-Hydroxy-dUMP and 5-Hydroxymethyl-dUMP in Thymidylate Synthase-Catalyzed Reaction in View of Molecular Modeling and Structural Studies", *Bioorganic Chemistry* 28: 156–162, 2000

[Kapetanovic, 2008] Kapetanovic IM. Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach. *Chem Biol Interact*. 2008 Jan 30;171(2):165-76.

[Kelly *et al.*, 2013] Kelly C, Bhuva N, Harrison M, Buckley A, Saunders M, "Use of raltitrexed as an altern ative to 5-fluorouracil and capecitabine in cancer patients with cardiac history", *European Journal of Cancer* 49, 2303-2310, 2013

[Kitchen *et al.*, 2004] Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov. 2004 Nov;3(11):935-49. Review.

[Koehn and Kohen, 2009] Koehn EM, Kohen A, "Flavin-dependent thymidylate synthase: A novel pathway towards thymine", *Archives of Biochemistry and Biophysics*, Vol. 493, Issue 1, pp. 96-102, 2009

[Lang *et al.*, 2009] Lang PT, Brozell SR, Mukherjee S, Pettersen EF, Meng EC, Thomas V, Rizzo RC, Case DA, James TL, Kuntz ID. DOCK 6: combining techniques to model RNA-small molecule complexes. *RNA*. 2009 Jun;15(6):1219-30.

[Lensink and Wodak, 2013] Lensink MF, Wodak SJ. Docking, scoring, and affinity prediction in CAPRI. *Proteins*. 2013 Dec;81(12):2082-95.

[Li *et al.*, 2001] Li Q, Boyer C, Lee JY, Shepard HM. A novel approach to thymidylate synthase as a target for cancer chemotherapy. *Mol Pharmacol*. 2001 Mar;59(3):446-52.

[Li *et al.*, 2014a] Li Y, Liu Z, Li J, Han L, Liu J, Zhao ZX, Wang R. Comparative Assessment of Scoring Functions on an Updated Benchmark: I. Compilation of the Test Set. *J Chem Inf Model*. 2014 Apr 9. [Epub ahead of print]

[Li *et al.*, 2014b] Li Y, Han L, Liu Z, Wang R. Comparative Assessment of Scoring Functions on an Updated Benchmark: II. Evaluation Methods and General Results. *J Chem Inf Model*. 2014 Apr 7. [Epub ahead of print]

[Liu *et al.*, 2013] Liu Y, Wu W, Hong W, Sun X, Wu J, Huang Q, „Raltitrexed-based chemotherapy for advanced colorectal cancer", *Clin Res Hepatol Gastroenterol.*, 2013

[Longley *et al.*, 2003] Longley DB, Harkin DP, Johnston PG, "5-Fluorouracil: mechanisms of action and clinical strategies", *Nature Reviews Cancer* 3, pp. 330-338, 2003

[Mader *et al.*, 1998] Mader RM, Müller M, Steger GG, "Review: Resistance to 5-Fluorouracil", *Gen. Pharmac*. Vol. 31, No. 5, pp. 661–666, 1998.

[Mas-Moruno *et al.*, 2007] Mas-Moruno C, Cruz LJ, Mora P, Francesch A, Messeguer A, Pérez-Paya E, Albericio F. Smallest peptoids with antiproliferative activity on human neoplastic cells. *J Med Chem.* 2007 May 17;50(10):2443-9.

[Merino *et al.*, 2010] Merino A, Bronowska AK, Jackson DB, Cahill DJ. Drug profiling: knowing where it hits. *Drug Discov Today*. 2010 Sep;15(17-18):749-56.

[Miller *et al.*, 2012] Miller BR, McGee TD, Swails JM, Homeyer N, Gohlke H, Roitberg AE. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. Journal of *Chemical Theory and Computation* 2012, 8 (9), 3314-3321

[Mirijanian *et al.*, 2014] Mirijanian DT, Mannige RV, Zuckermann RN, Whitelam S. Development and use of an atomistic CHARMM-based forcefield for peptoid simulation. *J Comput Chem.* 2014 Feb 15;35(5):360-70.

[Montfort and Weichsel, 1997] Montfort WR, Weichsel A, "Thymidylate Synthase: Structure, Inhibition, and Strained Conformations During Catalysis", *Pharmacol. Ther*. Vol. 76, Nos. 1-3, pp. 29-43, 1997

[Nandel and Saini, 2010] Nandel FS, Saini A. Peptoids with aliphatic sidechains as helical structures without hydrogen bonds and collagen/ inverse-collagen type structures. *Journal of Biophysical Chemistry* 2 (2011) 37-48

[Ntie-Kang *et al.*, 2013] Ntie-Kang F, Zofou D, Babiaka SB, Meudom R, Scharfe M, Lifongo LL, Mbah JA, Mbaze LM, Sippl W, Efange SM. AfroDb: a select highly potent and diverse natural product library from African medicinal plants. *PLoS One.* 2013 Oct 30;8(10):e78085

[Pettersen *et al.*, 2004] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera – a visualization system for exploratory research and analysis. *J Comput Chem.* 2004 Oct;25(13):1605-12.

[Phan *et al.*, 2001] Phan J, Koli S, Minor W, Dunlap RB, Berger SH, Lebioda L. Human thymidylate synthase is in the closed conformation when complexed with dUMP and raltitrexed, an antifolate drug. *Biochemistry*. 2001 Feb 20;40(7):1897-902.

[Piela, 2014] Piela L. Ideas of Quantum Chemistry. Second Edition. Elsevier, 2014

[Poltev *et al.*, 2008] Poltev VI, Anisimov VM, Danilov VI, Deriabina A, Gonzalez E, Jurkiewicz A, Leś A, Polteva N. DFT study of B-like conformations of deoxydinucleoside monophosphates containing Gua and/or Cyt and their complexes with Na+ cation. *J Biomol Struct Dyn*. 2008 Apr;25(5):563-71.

[Poltev *et al.*, 2009] Poltev VI, Anisimov VM, Danilov VI, Deriabina A, Gonzalez E, Garcia D, Rivas F, Jurkiewicz A, Leś A, PoltevaN, DFT study of minimal fragments of nucleic acid single chain for explication of sequence dependence of DNA duplex conformation, *Journal of Molecular Structure: THEOCHEM*, Vol. 912, Issues 1–3, 30 October 2009, pp. 53-59

[Ritchie, 2008] Ritchie DW. Recent progress and future directions in protein-protein docking. *Curr Protein Pept Sci*. 2008 Feb;9(1):1-15.

[Ruman *et al.*, 2007] Ruman T, Kuśnierz A, Jurkiewicz A, Leś A, Rode W. The synthesis, reactivity and 1H NMR investigation of the hydroxyborohydride anion. *Inorganic Chemistry Communications*, Volume 10, Issue 9, September **2007**, 1074-1078

[Ruman *et al.*, 2009a] Ruman T, Długopolska K, Kuśnierz A, Jurkiewicz A, Leś A, Rode W. "Synthesis and NMR properties of novel 5,6-dihydroborauracil derivatives", *Bioorganic Chemistry*, Volume 37, Issue 3, June **2009**, Pages 65-69

[Ruman *et al.*, 2009b] Tomasz Ruman, Karolina Długopolska, Dagmara Kramarz, Agata Jurkiewicz, Andrzej Leś, and Wojciech Rode, "The synthesis, reactivity and NMR investigation on 15N-thiophosphoramidates", *Letters in Organic Chemistry*, **2009**, vol 8(6) 642-647.

[Ruman *et al.*, 2010a] Ruman T, Długopolska K, Jurkiewicz A, Rydel K, Leś A, Rode W. The synthesis and NMR investigation on novel boron derivatives of stavudine. *Bioorg Chem.* 2010 Jun;38(3):87-91.

[Ruman *et al.*, 2010b] Ruman T, Długopolska K, Jurkiewicz A, Rut D, Fraczyk T, Cieśla J, Leś A, Szewczuk Z, Rode W. Thiophosphorylation of free amino acids and enzyme protein by thiophosphoramidate ions. *Bioorg Chem.* **2010** Apr;38(2):74-80.

[Salo-Ahen and Wade, 2011] Salo-Ahen OM, Wade RC. The active-inactive transition of human thymidylate synthase: targeted molecular dynamics simulations. *Proteins*. 2011 Oct;79(10):2886-99.

[Schaftenaar and Noordik, 2000] Schaftenaar G, Noordik JH. Molden: a pre- and post-processing program for molecular and electronic structures. *J Comput Aided Mol Des.* 2000 Feb;14(2):123-34.

[Stroud and Finer-Moore, 2003] Stroud RM, Finer-Moore JS. Conformational dynamics along an enzymatic reaction pathway: thymidylate synthase, “the movie”. *Biochemistry*. 2003 Jan 21;42(2):239-47

[Sui *et al.*, 2007] Sui Q, Borchardt D, Rabenstein DL. Kinetics and equilibria of cis/trans isomerization of backbone amide bonds in peptoids. *J Am Chem Soc*. 2007 Oct 3;129(39):12042-8

[Świniarska *et al.*, 2010] Swiniarska M, Leś A, Rode W, Cieśla J, Millán-Pacheco C, Blake IO, Pastor N. Segmental motions of rat thymidylate synthase leading to half-the-sites behavior. *Biopolymers*. 2010 Jun;93(6):549-59

[Valli *et al.*, 2013] Valli M, dos Santos RN, Figueira LD, Nakajima CH, Castro-Gamboa I, Andricopulo AD, Bolzani VS. Development of a natural products database from the biodiversity of Brazil. *J Nat Prod.* 2013 Mar 22;76(3):439-44.

[Voelz *et al.*, 2011] Voelz VA, Dill KA, Chorny I. Peptoid conformational free energy landscapes from implicit-solvent molecular simulations in AMBER. *Biopolymers*. 2011;96(5):639-50.

9 Appendices

1 Fusaro *et al*., 2010

Fusaro M, Jurkiewicz A, Jarmuła A, Leś A, Rode W, "Hypothesis of a proton switch in QM/MM modelling of interaction of dUMP analogues with thymidylate synthase", *Molecular Simulation* Volume 36, Issue 13, **2010**

2 Poltev *et al*., 2008

Poltev VI, Anisimov VM, Danilov VI, Deriabina A, Gonzalez E, Jurkiewicz A, Leś A, Polteva N. DFT study of B-like conformations of deoxydinucleoside monophosphates containing Gua and/or Cyt and their complexes with Na+ cation. *J Biomol Struct Dyn*. 2008 Apr;25(5):563-71.

3 Poltev *et al*., 2009

Poltev VI, Anisimov VM, Danilov VI, Deriabina A, Gonzalez E, Garcia D, Rivas F, Jurkiewicz A, Leś A, PoltevaN, DFT study of minimal fragments of nucleic acid single chain for explication of sequence dependence of DNA duplex conformation, *Journal of Molecular Structure: THEOCHEM*, Vol. 912, Issues 1–3, 30 October 2009, pp. 53-59

4 Ruman *et al*., 2009a

Ruman T, Długopolska K, Kuśnierz A, Jurkiewicz A, Leś A, Rode W. Synthesis and NMR properties of novel 5,6-dihydroborauracil derivatives. *Bioorganic Chemistry*, Volume 37, Issue 3, June **2009**, Pages 65-69

5 Ruman *et al*., 2007

Ruman T, Kuśnierz A, Jurkiewicz A, Leś A, Rode W. The synthesis, reactivity and 1H NMR investigation of the hydroxyborohydride anion. *Inorganic Chemistry Communications*, Volume 10, Issue 9, September **2007**, 1074-1078

6 Ruman *et al*., 2010a

Ruman T, Długopolska K, Jurkiewicz A, Rydel K, Leś A, Rode W. The synthesis and NMR investigation on novel boron derivatives of stavudine. *Bioorg Chem*. 2010 Jun;38(3):87-91.

7 Ruman *et al*., 2009b

Tomasz Ruman, Karolina Długopolska, Dagmara Kramarz, Agata Jurkiewicz, Andrzej Leś, and Wojciech Rode. The synthesis, reactivity and NMR investigation on 15N-thiophosphoramidates. *Letters in Organic Chemistry*, **2009**, vol 8(6) 642-647.

8 Ruman *et al*., 2010b

Ruman T, Długopolska K, Jurkiewicz A, Rut D, Fraczyk T, Cieśla J, Leś A, Szewczuk Z, Rode W. Thiophosphorylation of free amino acids and enzyme protein by thiophosphoramidate ions. *Bioorg Chem.* **2010** Apr;38(2):74-80.

9 Dąbrowska-Maś *et al*., 2012

Dąbrowska-Maś E, Frączyk T, Ruman T, Radziszewska K, Wilk P, Cieśla J, Zieliński Z, Jurkiewicz A, Gołos B, Wińska P, Wałajtys-Rode E, Leś A, Nizioł J, Jarmuła A, Stefanowicz P, Szewczuk Z, Rode W. Tyrosine nitration affects thymidylate synthase properties. *Org Biomol Chem.* 2012 Jan 14;10(2):323-31.

Molecular Simulation
Vol. 36, No. 13, November 2010, 1059–1066

Taylor & Francis
Taylor & Francis Group

Hypothesis of a proton switch in QM/MM modelling of interaction of dUMP analogues with thymidylate synthase

Massimo Fusaro[a]*, Agata Jurkiewicz[a], Adam Jarmuła[b], Andrzej Leś[a] and Wojciech Rode[b]

[a]*Faculty of Chemistry, Warsaw University, Pasteura 1, Warsaw, Poland;* [b]*Nencki Institute of Experimental Biology, Pasteura 3, Warsaw, Poland*

(*Received 26 February 2010; final version received 17 June 2010*)

Quantum mechanical, molecular mechanics and molecular dynamics (MD) methods were used to investigate initial steps of 2′-deoxyuridine-5′-monophosphate (dUMP) methylation catalysed by the thymidylate synthase (TS) enzyme. The amino acid residues surrounding the active site within a 10 Å radius sphere were modelled with the combined quantum mechanical (B3LYP/LANL2DZ) and molecular mechanics ONIOM double-layer method. The results indicated the initial nucleophilic attack of Cys146 on dUMP to be concerted with formation of a hydrogen bond to the oxygen O4 of dUMP. Moreover, the proton in the vicinity of the O4 atom appears to act as a 'proton switch': if a proton is present near O4, it stabilises the S(Cys146)—C6(dUMP) sulphur–carbon bond, but if it is absent, the sulphur–carbon bond does not form. If the O4 oxygen is replaced by sulphur atom, the 'switch effect' does not occur. The suggested correlation between the strength of hydrogen bond involving O4 oxygen and the ability of dUMP to form bonds at C6 corresponds well to the crystal structures of TS complexes available in the Protein Data Bank. In the vast majority of crystal structures, the presence of the S(Cys146)—C6(dUMP) bond was coupled with the presence of hydrogen bond between the dUMP O4 atom and the conserved Asn177. The 'proton switch' hypothesis is supported also by the results of MD studies of TS binary complexes, suggesting that average distance separating S(Cys146) and C6(dUMP) becomes distinctly shorter in the presence of hydrogen bonding between Asn177 and O4.

Keywords: thymidylate synthase; dUMP; methylenetetrahydrofolate; density functional theory

1. Introduction

Thymidylate synthase (TS, EC 2.1.1.45) is an enzyme catalysing the C(5) methylation of 2′-deoxyuridine-5′-monophosphate (dUMP) to 2′-deoxythymidine-5′-monophosphate with concerted conversion of the cofactor, methylenetetrahydrofolate (MTF), to dihydrofolate [1]. This reaction is the sole pathway of *de novo* thymine synthesis in a cell and is a target in chemotherapy [2–5], in particular therapy involving fluoro-pyrimidine drugs [6,7]. Basic structure of the ternary complex TS–dUMP–MTF is presented in Figure 1.

The reaction of TS catalysis is initiated by the nucleophilic addition of the active site cysteine (CYS 146) to the pyrimidine C(6) atom of dUMP. It is believed that at this stage the negative charge of cysteine residue is delocalised, probably towards the C(4)═O of dUMP, forming the corresponding enolate anion at the C(4)—O(4) moiety. It is believed that due to resonance between C(4)═O(4), C(4)—C(5) and C(5)═C(6) bonds, the C(5) atom becomes negative enough to allow for a consecutive attack of the methylene residue which takes the form of an iminium cation created through the opening of imidazolidine ring of the cofactor [8].

A strong TS inhibitor, 5-fluoro-dUMP (FdUMP), is the active form of 5-fluorouracil and 5-fluoro-2′-deoxyuridine – drugs used in antitumour chemotherapy. The inhibition mechanism involves time-dependent formation of a ternary, covalently bound complex of the enzyme with FdUMP and N5,10-MTF. As the complex is formed, the reaction stops, resulting in a slowly reversible inactivation of the enzyme. The substitution of the C(4)═O group in FdUMP pyrimidine ring by either C(4)═N—OH group (in N4-hydroxy-5-fluoro-2′-deoxycytidine-5′-monophosphate; N4-OH-FdCMP) or C(4)═S group (in 4-thio-FdUMP) preserves high inhibitory potency, but alters selectivity of inhibition [9–11].

It is assumed that dUMP pyrimidine C(4)═O and non-dissociated N(3)—H groups are vital for the specific binding of pyrimidine moiety by the enzyme through active site asparagine residue (Asn177) and an ordered water molecule [12,13]. The assumption is that the asparagine residue stabilises, by hydrogen bonding, the partial negative charge on O4 of covalently bound dUMP [14]. This suggests an explanation of TS active site discrimination between dUMP and dCMP. As it is shown, strong mechanism-based inactivation by FdUMP or 4-thio-FdUMP of TS depends on the fact that N(3)—H group

*Corresponding author. Email: massimo@tiger.chem.uw.edu.pl

ISSN 0892-7022 print/ISSN 1029-0435 online

DOI: 10.1080/08927022.2010.502938
http://www.informaworld.com

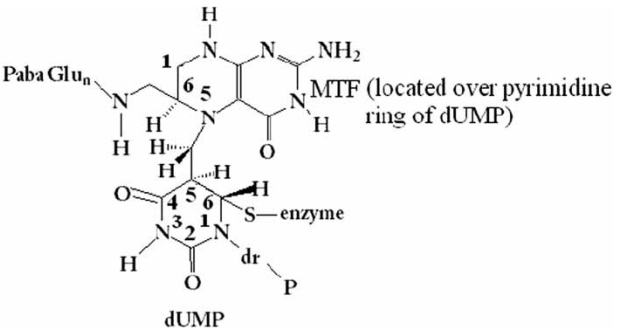

Figure 1. Schematic diagram of the TS–dUMP–MTF complex.

remains non-dissociated [15], hence it appears that the active site Asn177 is involved in the interaction. The same is probably true for N4—OH—FdCMP [8].

In order to learn more about the mechanism of interactions of dUMP and its 5-substituted analogues [FdUMP and 5-OH-dUMP (dUMP-OH)] with TS, and the role of 4-thio substitution in those interactions, the combined quantum mechanical and molecular mechanics modelling study was attempted and relevant data from molecular dynamics (MD) simulations of the structures of binary TS complexes with various other dUMP analogues, substituted at the C5 position in the base and/or in the C2′ position in the sugar, are presented. Findings from the computational work were compared with crystal structures of TS complexes available in the Protein Data Bank (PDB).

The numbering of residues of *Escherichia coli* TS is used in this work. The corresponding numbering of *Lactobacillus casei* TS is occasionally provided in parentheses.

The 'switch effect' hypothesis developed here can be summarised as follows: screening of the partial negative charge of O4 originating from the Cys146 attack on dUMP facilitates the formation of C6—S carbon–sulphur bond (so-to-say: no screening, no bond).

2. Method

2.1 *Models of the active site*

The active site of an enzyme contains the catalytic and binding sites. The structure and chemical properties of the active site allow for the recognition and binding of the substrate.

In other words, the active site is usually a small pocket at the surface of the enzyme, which contains residues responsible for the substrate specificity and catalytic residues that often act as proton donors or acceptors or are responsible for binding of a cofactor.

The X-ray diffraction structure of the monomer A of *E. coli* thymidylate synthase (ECTS) complexed with dUMP and MTF (PDB accession code 1KZI [16]) was used as the starting point for the calculations.

The hydrogen at C5 of dUMP was replaced either with fluorine, giving FdUMP, or with the hydroxyl group, giving dUMP-OH. The water molecules were removed from the X-ray structure.

Several models of different size that simulated the active site were considered. The small model, denoted as 5A model, was prepared by removing atoms farther than 5 Å from the S atom of CYS 146 residue, with the exception of atoms involved in non-peptidic bonds, which were included in the model. For the larger systems, denoted as 10A model, a similar procedure with the cut-off radius of 10 Å was adopted.

The residues included in our active site are consistent with the ones referred to in the literature [17]. In the 5A model of the active site, the following residues were included: TYR 94, LEU 143, ALA 144, PRO 145, CYS 146, HIS 147, ARG 166 and dUMP-X. The MTF cofactor was not included, as it is farther than 5 Å from the S atom of CYS 146. Moreover, to avoid unnatural expansion of the system, the border main chain atoms were frozen during the following optimisation at the experimental positions [18] (see Figure 2).

The 10A model contained the following residues: LEU 90, GLY 91, PRO 92, VAL 93, TYR 94, ALA 132, TRP 133, ASN 134,VAL 135, GLY 136, GLU 137, LEU 138, ASP 139, LYS 140, MET 141, ALA 142, LEU 143, ALA 144, PRO 145, CYS 146, HIS 147, ALA 148, PHE 149, LEU 163, TYR 164, GLN 165, ARG 166, SER 167, CYS 168, ASP 169, VAL 170, PHE 171, LEU 172, GLY 173,

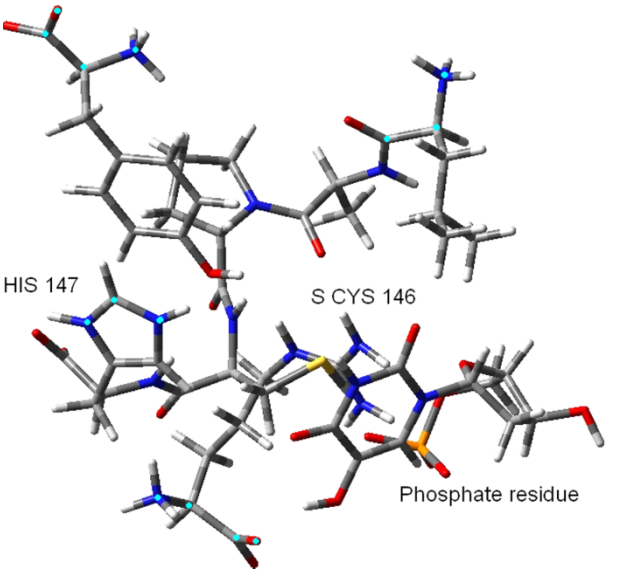

Figure 2. Schematic view of the 5A model. Fourteen atoms (C, N and O) indicated by cyan-coloured marks were frozen during the calculations to take the presence of the surrounding enzymatic environment into account (colour online).

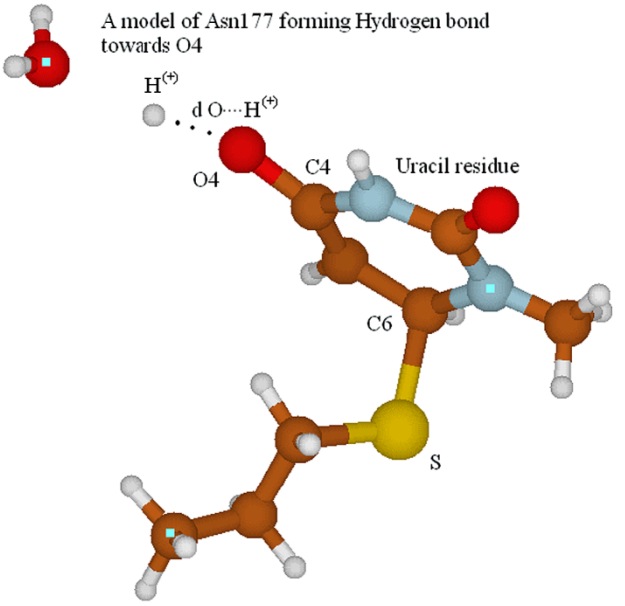

Figure 3. The S model of protonated residue approaching dUMP bond to CYS 146. The geometry corresponds to the transition state for proton movement. The atoms indicated by cyan-coloured marks, C, N and O of water, were frozen during the calculations.

LEU 174, PRO 175, PHE 176, ASN 177, ILE 178, ALA 179, SER 180, TYR 181, TRP 201, THR 202, GLY 203, GLY 204, ASP 205, THR 206, HIS 207, LEU 208, TYR 209, dUMP-X and MTF. In this case, also to avoid unnatural expansion of the system, the main chain atoms of the polypeptide were frozen in their experimental crystallographic positions, as in the work of Torrent et al. [18].

The dangling bonds in the models were saturated with hydrogen atoms.

The amino acid residues were modelled with the GaussView 3.07 [19] in their standard protonation states, adopting total charges of +1 (ARG, HIS and LYS), −1 (ASP and GLN) or 0 (all remaining amino acids).

The MTF cofactor had total charge of +1 and the phosphate residue had total charge of −2 ($—PO_4^{2-}$; 5A model) or −1 ($—HPO_4^-$; 10A model).

In the 10A and 5A models, a covalent bond exists between the S(Cys146) and C6 (dUMP) [16], and a proton is bonded to O4 atom of dUMP.

The resulting total charge of all models was neutral, since that allows for indirect simulation of the presence of counterions in the real systems. All models were prepared with the Chimera molecular visualisation program [20] and GaussView software [19].

In order to investigate the factors involved in the first step of the enzymatic reaction [17], i.e. the Michael addition of CYS 146 to the dUMP residue, a small S model was also created from a fragment of the 5A model, i.e. the coordinates of the atoms present in this small model are extracted from 1KZI PDB structure [16] (Figure 3). The S model simulates the approach of protonated residue (modelled as H_3O^+) towards the O4 oxygen of dUMP bound to Cys146 (modelled as 1-methyl uracil-6-mercaptopropane anion). The hydrogen atoms were added to saturate bonds and two carbon atoms were frozen during geometry optimisation to avoid unnatural expansion of the system.

2.2 *Calculations*

For the 5A and S models, the initial geometries based on literature data appropriately adapted were optimised with constraints using the quantum mechanical density functional theory with the B3LYP functional and the LANL2DZ [21] atomic basis set. The 5A models consist of 157 or 158 atoms (the latter in the case of dUMP-OH).

In order to incorporate protein environment effects to a larger extent, 10A models of the active site were also investigated. These models were subjected to the combined quantum mechanical/molecular mechanics (QM/MM) routine with mechanical embedding (two-layer ONIOM-ME) [22]. The internal part of the model (containing the same residues as in the 5A model) was investigated using the B3LYP/LANL2DZ method [18] (the QM-layer), whereas molecular mechanics method, employing the Amber force field [23], was applied for studying the external part (the MM-layer). In total, 149(150) and 737 atoms were included in the QM and MM layers, respectively.

All calculations were performed with the Gaussian 03 software [24]. The initial (before optimisation) and final (after optimisation) geometries were compared using the Tinker program [25].

3. Results

3.1 *Active site geometry*

The total QM/MM energies of the 5A and 10A models, as well as the mass-weighted, all-heavy-atom root mean square deviations (RMSD) between the initial and final geometries, are presented in Table 1.

Table 1 shows that the energy differences between the modelled systems depend almost entirely on the QM, while very little on the MM, parts; the former differing strongly from each other, while the latter agreeing very closely.

The RMSD between initial and final geometries of the 10A models are less affected during the optimisations than the corresponding RMSD in the 5A models due to more rigorous constraints imposed (freezing of main chain atoms in the 10A models vs. only border atoms in the 5A models).

Table 1. QM and MM energies and all-heavy-atom RMSD from the initial geometries in the 5A and 10A models.

	5A model		10A model	
System	Energy (Hartree) QM	RMSD (Å)	Energy (Hartree) QM/MM	RMSD (Å)
dUMP	−4048.1034	0.64	−3896.0053/91.9172	0.47
5-OH-dUMP	−4123.2206	0.86	−3971.2453/92.4893	0.48
FdUMP	−4147.3318	0.62	−3995.2899/92.0936	0.47

It was observed that during the geometry optimisation, when the proton connected to the oxygen O4 atom was removed (Figure 5), the distance separating the S (CYS 146) and C6 (dUMP-X) atoms lengthened considerably with respect to the initially assumed configuration (Figure 4). This result suggests that the formation of the S(Cys146)—C6(dUMP) bond is correlated with the presence of a hydrogen atom (or the formation of a strong hydrogen bond) at the O4 position of dUMP. Additional computational experiments showed that the S—C6 bond does not break if the O4 atom of dUMP is replaced with a sulphur atom or a thiol group. The S—C6 bond persists independently on the presence or absence of hydrogen bond at S4.

We can hypothesise that the hydrogen atom bound to the O4 atom of dUMP acts as a 'proton switch' stabilising the S(CYS 146)—C6 bond during the reaction. However, if the O4 atom is replaced with =S or —SH, the 'switch effect' is absent (presumably due to different charge distributions on the sulphur vs. oxygen atoms). The latter effect seems to contribute to the previously observed influence of the 4-thio substitution in FdUMP, consisting in a different specificity of slow-binding inhibition by 4-thio-FdUMP (of TSs from various sources) as compared to the specificity of inhibition by FdUMP [10].

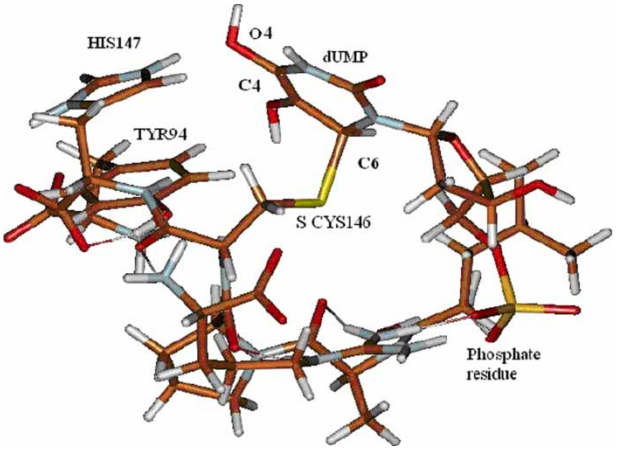

Figure 4. The final optimised structure of 5A model. The O4 atom of dUMP is protonated (zero total charge). The bond of CYS 146 S atom remains stable (1.8 Å as compared to 1.7 Å in 1KZI PDB file) during the geometry optimisation.

Additional calculations showed that the 'switch effect' occurs also in C(5)-substituted analogues of dUMP, i.e. when the hydrogen at the C5 position of dUMP is replaced by one of the following functional groups: —OH, —F, $-OCH_3$ or $-CH_2OH$. The ionisation state of the phosphate group ($-PO_4^{2-}$, $-HPO_4^-$ or $-H_2PO_4$) and consequently the total charge of the system (neutral system with proton bonded to the O4 atom and the phosphate group doubly ionised) have not influenced the 'switch effect'.

A transition state quantum mechanical calculation was performed, using the small S model, to show whether the presence of a proton close to the oxygen atom O4 can stabilise the S(CYS 146)—C6 bond, see Figure 3. The transition structure with only one imaginary frequency corresponding to the proton transfer from O4 atom to the water molecule was found at the O–H distance of 2.36 Å. Our simplified model corresponding to the vacuum rather than to biochemical systems does not contain either solvent or enzyme environment, hence the overestimation of calculated energy barrier of around 80 kcal/mol. A real barrier, however, must be considerably smaller as the proton transfer is driven by electrostatic forces which are screened by the molecular residues of the environment.

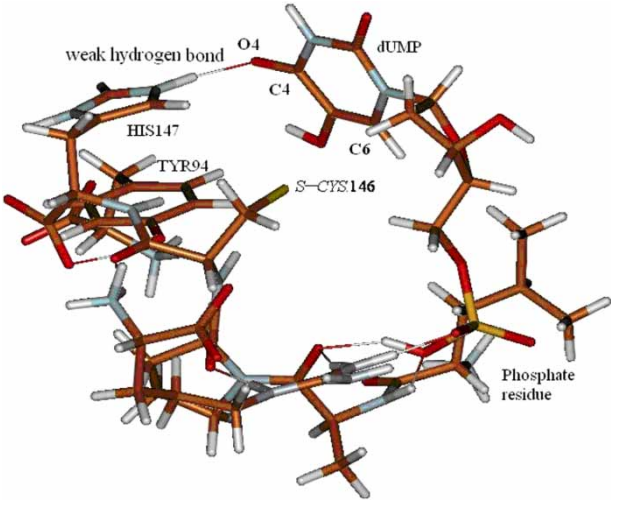

Figure 5. The final optimised structure of 5A model. The O4 atom of dUMP is not protonated. The bond of Cys146 S atom with C6 is not stable (>6 Å) during the geometry optimisation unless there is a proton connected to the O4 atom as shown in Figure 5. This is consistent with the reaction pathway described in [16]. The total charge of the model shown in the picture is zero.

Due to the presence of environment, the effective Coulombic forces are inversely proportional to the relative dielectric constant ε, which amounts to about 4 in the catalytic site interior in TS [4]. As a consequence, the forecast barrier should be reduced correspondingly. Moreover, experimental research shows [19] that the abstraction of the O4 proton in fact does take place. Therefore, the proton transfer can occur following the mechanism suggested here as the proton switch effect. The present simulation also indicates that the proton approaching dUMP residue bound to Cys146 can be an essential factor modulating the first step of the enzymatic reaction catalysed by TS.

3.2 Molecular dynamics

In the crystal structure of the ECTS–dUMP complex (PDB code 1BID) [26], the dUMP molecule is located away from Cys146 in such a way that the Michael addition to the pyrimidine C6 cannot occur. A distance of 3.16 Å observed in this structure can be perceived as a reference distance between the Cys146 S atom and the dUMP C6 atom in the absence of the cofactor. For the reaction to start, the cofactor binding must occur, upon which the active site closes and Cys146 is brought closer to dUMP, rendering the Michael addition feasible. An important factor that ensures a proper alignment of the cofactor at the time of binding is the orientation of the dUMP molecule, the correctness of which depends to a significant extent upon the presence of a double hydrogen bonding between the N3–H (donor) and O4 (acceptor) groups of dUMP and the active site asparagine (Asn177) [27–29].

Recently, MD simulations of the series of binary complexes between ECTS and 5- and/or 2′-substituted dUMP analogues were carried out in order to determine the effect of substitution on the conformation of the active site [30,31]. Considering our 'proton switch' hypothesis, particular attention should be paid to the variation of the C6–S distance in relation to the level of protonation of the O4 atom, the latter estimated in terms of preservation – or the lack thereof – of hydrogen bond between O4 and Asn177 as observed in the MD results.

3.3 Average structures from MD simulations

Average structures from the MD simulations can be divided into two groups depending on how much they conform to the active site organisation present in the crystal structure of the ECTS–dUMP complex. In general, a single C5-substitution in dUMP has a little (most of the substituents) to moderate ($-C_2H_5$ group) effect on the organisation of the active site [31], resulting in its strong preservation (first group). A similar effect takes place in the case of a double fluorine substitution, if one of the substituents lies at C5 such as in the 2′,5-diF-dUMP molecule [32]. The second group consists of the structures of three complexes, TS–5-C_3H_7-dUMP, TS–2′-F-*ara*-dUMP (fluorine in the 'up' configuration) and TS–2′,2″-diF-dUMP, wherein the substitution at dUMP C5 and/or C2′ results in the reorganisation of the active site [30,31] which leads to a considerable change in the positions with respect to each other of the catalytic CYS 146 and the pyrimidine ring of an analogue of dUMP, and in turn to an elongation of the molecular distance between S(CYS 146) and C6(dUMP). While the S–C6 distance in the structures from the first group fluctuates very little, ranging from 3.27 to 3.40 Å, it rapidly increases in the structures from the second group, from 3.75 Å in TS–2′-F-*ara*-dUMP to 4.36 Å in TS–2′,2″-diF-dUMP and to 5.65 Å in TS–5-C_3H_7-dUMP (Table 2). Apart from the S–C6 distance elongation, changes to the active site organisation manifest themselves in the TS–2′-F-*ara*-dUMP and TS–2′,2″-diF-dUMP structures in a displacement of the molecule of dUMP analogue from the position held by dUMP in the crystal structure of the TS–dUMP complex, which results in the uracil N3–H (TS–2′,2″-diF-dUMP) and O4 (TS–2′-F-*ara*-dUMP and TS–2′,2″-diF-dUMP) losing hydrogen bonding with Asn177 (Table 2). While in the TS–2′,2″-diF-dUMP structure, but not in TS–2′-F-*ara*-dUMP, the O4 atom gains a compensating hydrogen contact with the

Table 2. Distances separating C6(dUMP) and S(CYS 146) atoms and the presence (+) or absence (−) of hydrogen bonding in the average structures of binary complexes, as predicted by the results of MD simulations.

			Hydrogen contacts of Asn177 with	
Complexes from	Complex	C(6)–S(CYS 146) distance (Å)	N3—H	O4
First group	TS–5-X[a]-dUMP[b]	3.27–3.40	+	+
	TS–2′,5-diF-dUMP[c]	3.39	+	+
Second group	TS–2′-F-*ara*-dUMP[c]	3.75	+	−
	TS–2′,2″-diF-dUMP[c]	4.36	−	−
	TS–5-C_3H_7-dUMP[a]	5.65	+	+

[a] X = —F, —Cl, —Br, —I, —CHO, —CH_3, —C_2H_5, or CH_2OH. [b] Data from Ref. [26]. [c] Data from Ref. [27].

hydroxyl of Tyr94; dislocation of the molecule of 2′,2″-diF-dUMP prevents its C6 atom from maintaining an approachable distance to Cys146.

The structure of the TS–5-C_3H_7-dUMP complex has a special active site arrangement among the members of the second group [28]. In this structure, the large size of the propyl moiety resulted in the pyrimidine ring twisting and adopting the *syn* orientation relative to the sugar moiety (different from the usual *anti* conformation present in the rest of simulated structures as well as in the ECTS–dUMP crystal structure) in order to dispose the propyl between the indole rings of the two tryptophanes, Trp80 and Trp83, pushing them further away from each other as compared to their 'normal' positions. This rearrangement strongly dislocated the position of CYS 146, but did not disturb the hydrogen bonding of the uracil O4 and N3–H groups with Asn177 (Table 2).

An additional argument supporting the hypothesis of 'switch effect' on the C(4)═O residue comes from the analysis of the Asn177–O4 distance during MD simulation of the TS dimer complexed with dUMP/FdUMP and the cofactor. The results are presented in Figure 6 (TS with dUMP and cofactor) and Figure 7 (TS with FdUMP and cofactor). The variation of the Asn177–O4 distance in both systems clearly shows high flexibility of proton location in the vicinity of the O4 oxygen. Thus, it may be inferred that the proton can move from Asn177 towards dUMP/FdUMP and temporarily forms a hydrogen bond.

In general, the results of MD remain in line with the 'proton switch' hypothesis by showing the S(CYS 146)—C6(dUMP analogue) molecular distance to become longer and thus less approachable (in view of the chances for the formation of covalent bond) upon the lack of hydrogen bonding between the O4 atom of dUMP and Asn177 residue of ECTS, whereas shorter and more approachable upon the presence thereof. However, the results suggest that the hydrogen bond at O4 can be a necessary but not sufficient condition for a successful CYS 146 approach to the C6 atom. The active site organisation similar to the one observed in the crystal structure of the native TS–dUMP complex seems a prerequisite for the latter to take effect.

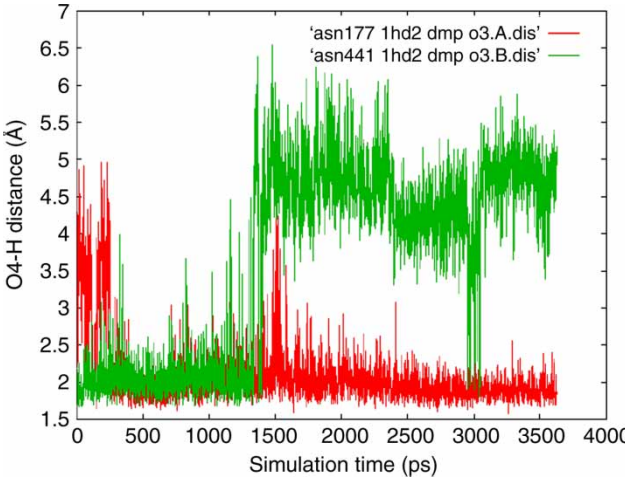

Figure 6. Distances (Å) between H atom of Asn177 (denoted as Asn441 in monomer B) and O4 atom of dUMP (denoted here as O3) for both monomers of TS, registered during 3 ns MD simulation of TS–dUMP–MTF ternary complex. Red, monomer A; green, monomer B (colour online).

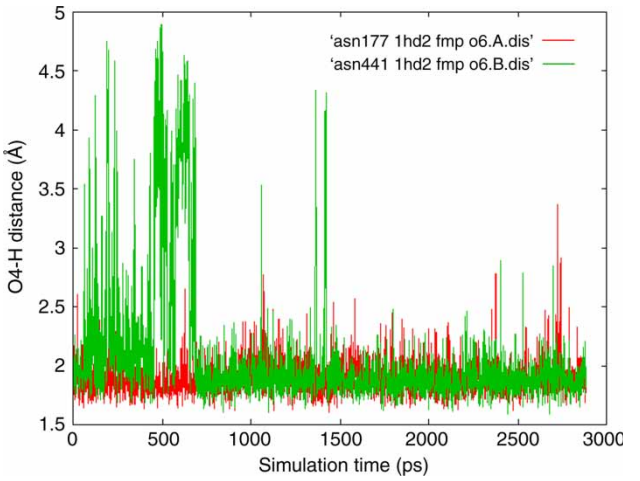

Figure 7. Distances (Å) between H atom of Asn177 (denoted as Asn441 in monomer B) and O4 atom of dUMP (denoted here as O6) for both monomers of TS, registered during 3 ns MD simulation of TS–FdUMP–MTF ternary complex. Red, monomer A; green, monomer B (colour online).

3.4 Crystal structures from PDB

Inspection of crystal structures of TS complexes stored in the PDB data bank shows that the presence of the S(Cys146)—C6(dUMP) covalent bond is correlated, in vast majority of cases, with the hydrogen bonding at the O4 atom. In most cases, one H-bond donated to O4 comes from Asn177, an essential amino acid residue aiding in recognising and orienting the molecule of dUMP. Additional hydrogen contact of O4 with Wat 1, a conserved water molecule participating in a water-mediated hydrogen bond network coordinated by glutamate 58, is usually also present. Only two structures were found, in which the H-bond between O4 and Asn177 is not present; the electron densities as well as distance and angular conditions, however, suggest the presence of the S—C6 bond. Both these structures are TS mutants: E58(60)Q from *L. casei* (LCTS) in complex with dUMP and a cofactor analogue CB3717 (PDB accession code 1VZE) [32], and N177A from *E. coli* (ECTS) in an analogous complex with dUMP and CB3717 (PDB accession code 1BQ1) [33]. In the structure of the LCTS–dUMP–CB3717 complex, dUMP shifted farther from its original position, with the O4 atom loosing the

hydrogen bonding with Asn177(229) and Wat 1, while forming a hydrogen bridge through another water molecule (Wat C) with Tyr94(146). In the structure of the ECTS–dUMP–CB3717 complex, on the other hand, asparagine 177 was mutated into a smaller residue of alanine, partly freeing a space occupied in the wild-type enzyme by the side chain of Asn177 and resulting in the side chain of His147 relocating into that space and forming a compensating hydrogen bond with O4. Another H-bond, linking O4 with the Wat C molecule, was also present.

Thus, in view of the crystal structures of TS complexes, hydrogen bonding to dUMP O4 in the form of direct bonding from the enzyme (Asn177) and/or an indirect bond, mediated by water, appears as a key factor for the presence of a covalent linkage between C6(dUMP) and S(CYS 146).

4. Conclusions

The nucleophilic attack of CYS 146 on the C6 carbon atom of dUMP appears to be concerted with the bonding of a proton to the oxygen O4 atom of dUMP or with the presence of a proton donor forming hydrogen bond with the O4 atom. Moreover, the proton bound to the O4 atom seems to act as a 'proton switch' which stabilises or destabilises the sulphur–carbon S(Cys146)—C6(dUMP) bond. The 'switch effect' does not occur when the O4 atom is replaced with a thione or thiol sulphur.

Acknowledgements

The computational work was performed on the computers of the Interdisciplinary Centre for Mathematical and Computational Modelling, Warsaw University.

References

[1] C.W. Carreras and D. Santi, *The catalytic mechanism and structure of thymidylate synthase*, Annu. Rev. Biochem. 64 (1995), pp. 721–762.

[2] P.V. Danenberg, *Thymidylate synthetase - a target enzyme in cancer chemotherapy*, Biochim. Biophys. Acta 473 (1977), pp. 73–92.

[3] L.W. Hardy, J.S. Finer-Moore, W.R. Montfort, M.O. Jones, D.V. Santi, and R.M. Stroud, *Atomic structure of thymidylate synthase: target for rational drug design*, Science 235 (1987), pp. 448–455.

[4] B.K. Shoichet, R.M. Stroud, D.V. Santi, I.D. Kuntz, and K.M. Perry, *Structure-based discovery of inhibitors of thymidylate synthase*, Science 259 (1993), pp. 1445–1450.

[5] Y. Takemura and A.L. Jackman, *Folate-based thymidylate synthase inhibitors in cancer chemotherapy*, Anti-Cancer Drugs 8 (1997), pp. 3–16.

[6] A.L. Parr, T.G. Myers, S.L. Holbeck, Y.J. Loh, and C.J. Allegra, *Thymidylate synthase as a molecular target for drug discovery using the National Cancer Institute's Anticancer Drug Screen*, Anti-Cancer Drugs 12 (2001), pp. 569–574.

[7] G.J. Peters, H.H.J. Backus, S. Freementle, B. van Triest, G. Codacci-Pisanelli, C.L. van der Wilt, K. Smid, J. Lunec, A.H. Calvert, S. Marsh, H.L. McLeod, E. Bloemena, S. Meijer, G. Jansen, C.J. van Groeningen, and H.M. Pinedo, *Induction of thymidylate synthase as a 5-fluorouracil resistance mechanism*, Biochim. Biophys. Acta 1587 (2002), pp. 194–205.

[8] W. Rode and A. Leś, *Molecular mechanism of thymidylate synthase-catalyzed reaction and interaction of the enzyme with 2- and/or 4-substituted analogues of dUMP and 5-fluoro-dUMP*, Acta Biochim. Pol. 43 (1996), pp. 133–142.

[9] W. Rode, Z. Zieliński, J.M. Dzik, T. Kulikowski, M. Bretner, B. Kierdaszuk, J. Cieśla, and D. Shugar, *Mechanism of inhibition of mammalian tumor and other thymidylate synthases by N4-hydroxy-dCMP, N4-hydroxy-5-fluoro-dCMP, and related analogues*, Biochemistry 29 (1990), pp. 10835–10842.

[10] J.M. Dzik, Z. Zieliński, J. Cieśla, M. Bretner, T. Kulikowski, D. Shugar, J.R. Bertino, and W. Rode, *Interaction of 2-thio-5-fluoro-dUMP and 4-thio-5-fluoro-dUMP with mammalian normal and tumour and helminthic thymidylate synthases: influence of C(4)-substituents on specificity for enzyme inactivation*, Biochem. Biophys. Res. Commun. 195 (1993), pp. 1301–1308.

[11] K. Felczak, A. Miazga, J. Poznański, M. Bretner, T. Kulikowski, J.M. Dzik, B. Gołos, Z. Zieliński, J. Cieśla, and W. Rode, *5-Substituted N(4)-hydroxy-2'-deoxycytidines and their 5'-monophosphates: synthesis, conformation, interaction with tumor thymidylate synthase, and in vitro antitumor activity*, J. Med. Chem. 43 (2000), pp. 4647–4656.

[12] L.W. Hardy and E. Nalivaika, *Asn177 in Escherichia coli thymidylate synthase is a major determinant of pyrimidine specificity*, Proc. Natl Acad. Sci. USA 89 (1992), pp. 9725–9729.

[13] L. Liu and D.V. Santi, *Mutation of asparagine 229 to aspartate in thymidylate synthase converts the enzyme to a deoxycytidylate methylase*, Biochemistry 31 (1992), pp. 5100–5104.

[14] D.A. Matthews, J.E. Villafranca, C.A. Janson, W.W. Smith, K. Welsh, and S. Freer, *Stereochemical mechanism of action for thymidylate synthase based on the X-ray structure of the covalent inhibitory ternary complex with 5-fluoro-2'-deoxyuridylate and 5,10-methylenetetrahydrofolate*, J. Mol. Biol. 214 (1990), pp. 937–948.

[15] J.M. Dzik, T. Kulikowski, Z. Zieliński, J. Cieśla, W. Rode, and D. Shugar, *Interaction of 5-fluoro-4-thio-2'-deoxyuridine 5'-phosphate with mammalian tumour thymidylate synthase: role of the pyrimidine N(3)-H dissociation*, Biochem. Biophys. Res. Commun. 149 (1987), pp. 1200–1207.

[16] T.A. Fritz, L. Liu, J.S. Finer-Moore, and R.M. Stroud, *Tryptophan 80 and leucine 143 are critical for the hydride transfer step of thymidylate synthase by controlling active site access*, Biochemistry 41 (2002), ppp. 7021–7029.

[17] J.S. Finer-Moore, D.V. Santi, and R.M. Stroud, *Lessons and conclusions from dissecting the mechanism of a bisubstrate enzyme: thymidylate synthase mutagenesis, function, and structure*, Biochemistry 42 (2003), pp. 248–256.

[18] M. Torrent, T. Vreven, D.G. Musaev, K. Morokuma, O. Farkas, and H.B. Schlegel, *Effects of the protein environment on the structure and energetics of active sites of metalloenzymes. ONIOM study of methane monooxygenase and ribonucleotide reductase*, J. Am. Chem. Soc. 124 (2002), pp. 192–193.

[19] R. Dennington, II, T. Keith, J. Millam, K. Eppinnett, W.L. Hovell, and R. Gilliland, *GaussView Version 3.09*, Semichem, Inc., Shawnee Mission, KS, 2003.

[20] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, and T.E. Ferrin, *UCSF Chimera - A visualization system for exploratory research and analysis*, J. Comput. Chem. 25 (2004), pp. 94143–2240.

[21] P.J. Hay and W.R. Wadt, Ab initio *effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg*, J. Chem. Phys. 82 (1985), ppp. 270–283.

[22] T. Vreven, K. Morukuma, O. Farkas, H.B. Schlegel, and M.J. Frish, *Geometry optimization with QM/MM, ONIOM, and other combined methods. I. Microiterations and constraints*, J. Comput. Chem. 24 (2003), ppp. 760–769.

[23] Yong Duan, Chun Wu, Shibasish Chowdhury, Mathew C. Lee, Guoming Xiong, Wei Zhang, Rong Yang, Piotr Cieplak, Ray Luo, Taisung Lee, James Caldwell, Junmei Wang, Peter Kollman. *A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations*, Journal of Computational Chemistry, 24, 16, 1999–2012.

[24] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery, Jr, T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, and J.A. Pople, *Gaussian 03, Revision C.01*, Gaussian, Inc., Wallingford, CT, 2004.

[25] P. Ren and J.W. Ponder, *Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation*, J. Phys. Chem. B 107 (2003), pp. 5933–5947.

[26] T.J. Stout, C.R. Sage, and R.M. Stroud, *The additivity of substrate fragments in enzyme-ligand binding*, Structure 6 (1998), pp. 839–848.

[27] R.M. Stroud and J.S. Finer-Moore, *Conformational dynamics along an enzymatic reaction pathway: thymidylate synthase, "the movie"*, Biochemistry 42 (2003), pp. 239–247.

[28] L. Liu and D.V. Santi, *Exclusion of 2′-deoxycytidine 5′-monophosphate by asparagine 229 of thymidylate synthase*, Biochemistry 32 (1993), pp. 9263–9267.

[29] J.S. Finer-Moore, L. Liu, C.E. Schafmeister, D.L. Birdsall, T. Mau, D.V. Santi, and R.M. Stroud, *Partitioning roles of side chains in affinity, orientation, and catalysis with structures for mutant complexes: asparagine-229 in thymidylate synthase*, Biochemistry 35 (1996), pp. 5125–5136.

[30] A. Jarmuła, P. Cieplak, T.M. Krygowski, and W. Rode, *The effect of 5-substitution in the pyrimidine ring of dUMP on the interaction with thymidylate synthase: molecular modeling and QSAR*, Bioorg. Med. Chem. 15 (2007), pp. 2346–2358.

[31] A. Jarmuła, A. Dowierciał, and W. Rode, *A molecular modeling study of the interaction of 2′-fluoro-substituted analogues of dUMP/FdUMP with thymidylate synthase*, Bioorg. Med. Chem. Lett. 18 (2008), pp. 2701–2708.

[32] D.L. Birdsall, W. Huang, D.V. Santi, R.M. Stroud, and J.S. Finer-Moore, *The separate effects of E60Q in Lactobacillus casei thymidylate synthase delineate between mechanisms for formation of intermediates in catalysis*, Protein Eng. 11 (1998), pp. 171–183.

[33] C.L. Reyes, C.R. Sage, E.E. Rutenber, R.M. Nissen, J.S. Finer-Moore, and R.M. Stroud, *Inactivity of N229A thymidylate synthase due to water-mediated effects: isolating a late stage in methyl transfer*, J. Mol. Biol. 284 (1998), pp. 699–712.

Journal of Biomolecular Structure & Dynamics, ISSN 0739-1102
Volume 25, Issue Number 5, (2008)

DFT Study of B-like Conformations of Deoxydinucleoside Monophosphates Containing Gua and/or Cyt and their Complexes with Na^+ Cation

http://www.jbsdonline.com

Valery I. Poltev[1,2,*]
Victor M. Anisimov[3]
Victor I. Danilov[4]
Alexandra Deriabina[1]
Eduardo Gonzalez[1]
Agata Jurkiewicz[5]
Andrzej Leś[5]
Nina Polteva[2]

[1]Benemerita Universidad Autonoma de Puebla
Puebla, 72570, Mexico
[2]Institute of Theoretical
and Experimental Biophysics
Russian Academy of Sciences
Pushchino, Moscow Region
142290, Russia
[3]Department of Pharmaceutical Sciences
School of Pharmacy
University of Maryland
20 Penn Street
Baltimore, MD 21201, USA
[4]Institute of Molecular Biology and Genetics
National Academy of Sciences of Ukraine
150 Zabolotny Street
Kyiv-143, 03143, Ukraine
[5]Department of Chemistry
University of Warsaw
Pasteura 1, 02-093 Warsaw, Poland

[*]Email: vpoltev@yahoo.com

Abstract

B-like minimum energy conformations of deoxydinucleoside monophosphate anions (dDMPs) containing Gua and/or Cyt and their Na^+ complexes have been studied by the DFT PW91PW91/DZVP method. The optimized geometry of the dDMPs is in close agreement with experimental observations and the obtained minimum energy conformations are consistent with purine-purine, purine-pyrimidine, and pyrimidine-purine arrangements in crystals of B-DNA duplexes. All the studied systems are characterized by pyramidalization of the amino groups, which participate in the formation of unusual hydrogen bond between the carbonyl oxygen of the second base in the dGpdC, dCpdG dDMPs, and their Na^+ complexes. In all the obtained structures the bases assume a nearly parallel disposition to each other and this effect is independent on the degree of their spatial superposition. From this it is concluded that the parallel disposition of the bases in the B-like single-stranded conformations is dictated by the sugar-phosphate backbone. Correspondingly, the base-base interactions attain a secondary role in the formation of these spatial structures. The formation of a weak C6-H6…O5′ hydrogen bond between cytosine and the phosphate oxygen is reported, in agreement with experimental observations.

Key words: dDMP; Dinucleoside; Deoxydinucleoside; Gua; Cyt; DFT; and Conformation.

Introduction

Computer simulation of nucleic acids is an indispensable tool for understanding molecular mechanisms of genetic processes. In spite of the enormous success achieved in this field by using modern molecular mechanical methods [see, *e.g.*, Ref. (1)], studying novel and underlying structural effects in these systems [see, *e.g.*, Ref. (2-4)] requires the use of sophisticated quantum mechanical methods. In the end, the knowledge gained from such studies will be a valuable aid in developing more accurate force fields.

During the last decade, several density functional theory (DFT) and correlated *ab initio* calculations were published on nucleic acid bases, their pair combinations, nucleosides and nucleotides [see Ref. (2, 3, 5-7) and references therein]. Despite their usefulness these publications were focused on relatively small building blocks of DNA and RNA. But in order to develop predictive theoretical models of DNA it is necessary to go to larger fragments of nucleic acids. Due to the exponential growth of computational cost with the number of atoms in the system such computations still represent a challenge. This paper represents a step toward the goal of studying larger systems and presents the results of DFT computations of minimum energy conformations of deoxydinucleoside monophospate anions (dDMPs) and their complexes with a Na^+ cation. In the present study, energy minimum B-conformations of dDMPs anions and their Na^+ complexes containing

guanine and/or cytosine nucleosides are considered, which are the largest structural units of DNA so far studied by using correlated quantum mechanical (QM) methods. A part of the results was presented at the 15th Albany Conversation on Biomolecular Stereodynamics (8).

Methods

To study dDMPs and their Na^+ complexes we used the DGauss program (9). The intrinsic feature of this program is reducing four-center Coulomb integrals to three-center ones by introducing the so-called fitting basis set. In this approximation, the Coulomb energy contains only three-center terms, making feasible QM study of relatively large systems. We used split-valence double zeta basis set (DZV), which is specifically optimized for DFT calculations (9, 10) as opposed to the commonly used basis sets, such as Pople's 6-31G(d) (11), that are optimized for Hartree-Fock calculations. DZV uses a double zeta basis for the chemically important valence region, while using a single function to represent the less mobile core orbitals. The calculations performed in this work employed the DZVP basis set augmented with polarization functions (10).

Energy values and molecular properties were calculated using the original gradient-corrected exchange-correlation functional developed by Perdew and Wang (12). It appears that the functional PW91PW91, based on the results of Kurita for dimers of rare gases (13), describes van der Waals forces better than commonly used functionals of Becke type (14). According to Kurita *et al*. (13), the Becke-type DFT functionals completely fail to describe the binding states, while the Perdew-Wang exchange-correlation functional provided a qualitatively correct trend in energy and geometry of the rare gas dimers. This conclusion is additionally supported by the recent DFT study of stacking complexes of DNA bases (15) where the PW91PW91 functional provided an improved description of these complexes in comparison with BLYP, BP86, OLYP, B3LYP, and X3LYP ones. In addition to that, the results of our study (see below) demonstrate that the properties of the systems considered are governed mainly by properties of the sugar-phosphate backbone rather than by base stacking.

To test the sensitivity of the obtained results to the choice of a particular DFT functionals, the basis sets, and to the density fitting, additional minimizations of selected structures were performed using the 6-31G(d,p) basis set (11), and the B3LYP functional (16-18) as implemented in the GAMESS program (19) and in the GAUSSIAN 03 program package (20). The structures obtained had the conformational parameters only slightly deviating from those of the initial ones, thus confirming that the conclusions of this work are invariant of the details in the DFT method.

Starting B-like geometries of dDMPs were generated using the HyperChem program version 7.01 (HyperCube, Inc.). Due to limitations in the HyperChem software the sugar rings were constructed with C3′-exo (not C2′-endo) puckering, characterized by a pseudorotation phase angle P = 192°. The possibility of biasing the computational results toward a particular initial puckering form is ruled out by a wide variety of different phase angles obtained after geometry optimization. This proves that the computational results are not particularly sensitive to initial sugar ring puckering in the region around the C2′-endo conformation. Neutral dDMPs were obtained by placing sodium cations at equidistant positions to the anionic oxygens (at O…Na distance of 2.2 Å) in the O1P-P-O2P plane. The generated structures were subjected to all-atom geometry optimization and the geometric characteristics of the optimized minimum energy conformations including bond lengths, valence and torsion angles as well as hydrogen-bond characteristics and distances from Na^+ ion to anion oxygens were analyzed. The program PROSIT (21) was used for the analysis of sugar-ring puckering and glycosidic torsions.

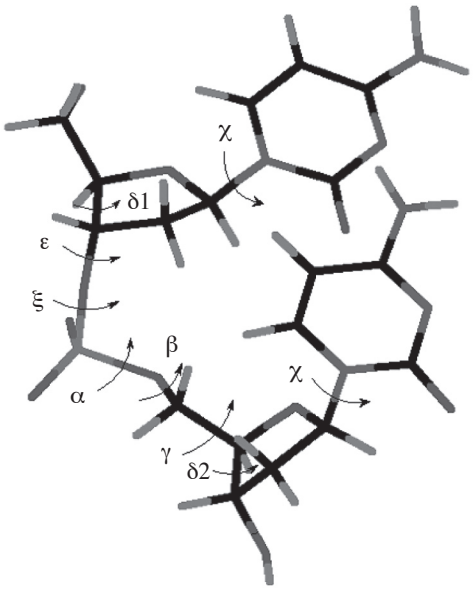

Figure 1: Designation of the torsion angles of the sugar-phosphate backbone of the dDMPs and their Na^+ complexes.

General Features of Minimum Energy Conformations. Sugar-phosphate Geometry and Na^+ Ion Positions

In the energy minimized structures the dDMPs and their Na^+ complexes adopt distorted B-like conformations with the torsion angles of the backbone and the mutual positions of the bases being characteristic of B-family DNA conformations. The computed bond lengths and valence angles are within the limits of the experimentally observed values (22, 23) with the exception of those of the anionic phosphate group, which will be discussed further. The Na^+ ion is located at an equidistant position of 2.26 Å from the anionic O1P and O2P oxygen atoms staying practically in the O1P-P-O2P plane.

Table I
Torsion Angles of the Sugar-Phosphate Backbone (definitions are in Figure 1) and Average Values of the Angles between the Base Planes (ϕ) for the Calculated Conformations of dDMPs and their Na^+ Complexes.

dDMP	δ1	ε	ζ	α	β	γ	δ2	ϕ
dCpdC	138.2	181.8	263.5	292.1	197.6	55.3	139.2	8.5
dCpdCNa^+	138.9	181.7	261.2	295.0	183.3	59.2	132.1	4.9
dGpdG	140.9	155.4	277.6	274.5	245.6	57.2	148.3	18.2
dGpdGNa^+	141.7	173.9	259.2	298.0	190.7	54.5	141.8	14.1
dCpdG	138.5	177.7	268.1	290.1	203.6	49.9	144.0	18.5
dCpdGNa^+	138.1	179.7	265.2	296.9	184.7	53.1	134.9	21.2
dGpdC	128.4	176.3	264.2	290.3	214.0	52.7	144.0	9.4
dGpdCNa^+	136.8	182.0	260.6	298.8	183.6	55.3	134.6	11.9
SD[a]	8.7	41.2	20.8	1.4	31.7	1.3	1.3	
SD[b]	6.3	34.7	27.6	11.9	2.1	3.0	9.4	
Exp.[c]	145.2	214.0	289.2	285.3	183.5	52.5	145.2	
Exp. SD[c]	4.9	8.6	4.8	9.8	13.0	5.7	4.9	

[a]Standard deviation for computed anionic dDMPs; [b]Standard deviation for computed sodium complexes; [c]Experimental data from Ref. (22).

The analysis of the computational data on the sugar-phosphate backbone is presented in Table I along with standard deviations from experimental data on dinucleoside monophosphates (DMP) and trinucleoside diphosphates (TDP) (23). Each of the torsion angles stays in the limits of one of the regions observed for DMPs and TDPs in crystals (23), namely, *-sc* (*-gauche*) for α and ζ, *ap* (*trans*) for β and ε, *+sc* (*+gauche*) for γ; δ values are characteristic for C2′-endo sugar puckering and clustered around 140°. The largest deviations from the experimental values are observed for the dGpdG anion. The RMS deviations from experiment for the calculated structures are smaller for the Na^+ complexes than those for the anions. This is quite reasonable because crystals carry net zero charge.

Geometry of Phosphate Group

Among all structural elements of the dDMPs the phosphate group shows the largest geometrical differences between the computed and experimental data (Table II). For the anionic dDMPs the P-O3′ and P-O5′ bonds are longer by 0.1 Å than the experimental value of 1.60 Å; the computed O1P-P-O2P angles overestimate the experimental value of 119° by 8°; and the O3′-P-O5′ angles are predicted to be smaller by 4° than the corresponding experimental value of 104°. The formation of a complex with Na^+ considerably improves the agreement with the experimental data giving 1.65 Å, 115° and 104°, respectively, for the corresponding geometric parameters. This is consistent with the notion that the neutralized dDMPs represent better models of the crystals than the anionic dDMPs.

From the analysis one may see that the experimental value for the O1P-P-O2P (O=P=O) angle is located between the calculated values for anionic and Na^+ com-

plexes. While this may be considered as merely a reasonable agreement between experiment and theory, and reserving room for possible limitations in the selected level of theory (24), it may also be viewed as the consequence of intermolecular interactions with water molecules in the crystals, which shifts Na^+ away from its position corresponding to the energy minima for the dDMP-Na^+ complexes, thus making a real system partially resembling the dDMP anion. The character of the differences upon going from the anionic to the sodium complex of the dDMPs is also reproduced by the B3LYP/6-31G(d,p) method.

To clarify the influence of the level of theory on the geometry of the phosphate group we performed calculations of dimethylphosphate (DP) in anionic form and in a complex with sodium using the MP2 method (25, 26). The comparison of the computation results on bond distances and valence angles pertaining to the phosphate group is presented in Table II. The MP2 results basically repeat the pattern in the geometrical properties of neutralized and anionic forms of the phosphate group discussed above, confirming the suitability of the DFT level. Surprisingly the better agreement with experiment is obtained with somewhat moderate 6-31+G(d) basis set (27) (possibly due to fortuitous error cancellation) and with the aug-cc-pVTZ one, the most extended basis set used here (28, 29). In order to improve the agreement for the anions further, higher level calculations are warranted [see, *e.g.*, calculations for anions of uracil dimers and trimers Ref. (30)]. However, good agreement observed for the Na^+ complexes justifies the selected methodology.

Table II
Phosphate Group Geometry at Various Levels of Theory.

Level of theory	System	P=O	P-O[a]	O=P=O[b]	O3'-P-O5'
B3LYP/6-31+G(d)	neutralized DP	1.523	1.632	113.7	103.7
B3LYP/6-31+G(d)	anionic DP	1.504	1.683	125.5	99.5
MP2/6-31+G(d)	neutralized DP	1.527	1.635	114.9	102.9
MP2/6-31+G(d)	anionic DP	1.509	1.684	125.9	98.9
MP2/aug-cc-pVDZ	neutralized DP	1.547	1.657	115.3	102.5
MP2/aug-cc-pVDZ	anionic DP	1.529	1.706	126.8	98.5
MP2/aug-cc-pVTZ	neutralized DP	1.515	1.618	115.3	103.1
MP2/aug-cc-pVTZ	anionic DP	1.498	1.663	125.5	99.4
DGauss/DZVP	neutralized dDMPs	1.53	1.65	115	104
DGauss/DZVP	anionic dDMPs	1.51	1.71	127	100
Experiment [c]	DMP & TDP	1.485	1.60	119	104

[a]Average values of the O3'-P and O5'-P distances; [b]Same as O1P-P-O2P; [c]Ref. (22); The geometry of DP is minimized at the gauche-gauche geometry of the methoxy groups.

Sugar Ring Puckering and Sugar-base Orientation

Analysis of the sugar ring geometry is performed using the terminology of phase angle, P, and puckering amplitude, υ_{max}, proposed by Altona and Sundaralingam (31). All the sugars of dDMP-Na^+ complexes have pseudorotational phase angles P laying in between 150º and 177º, *i.e.*, they pertain to C2′-endo conformations, with the sugar puckering amplitude being in the region between 31.2º and 35.3º. The sugar puckering of anionic dDMP varies in somewhat broader region with P being in the range from 136º to 206º, corresponding to C1′-exo, C2′-endo, or C3′-exo sugar types. Numerical characteristics of sugar rings and glycosyl torsion angles of dDMPs, dDMP-Na^+ complexes, and individual deoxynucleosides are listed in Table III.

Sugar ring bond lengths, valence angles, and endo-cyclic torsions of the calculated conformations are close to the averaged crystal structure values (22). Detailed comparison of the calculated endo-cyclic bond angles with crystal data is provided in Table IV.

The mutual orientations of the sugar and base rings correspond to the anti region of the glycosyl torsion angle χ (O4′-C1′-N9-C4 for guanine and O4′-C1′-N1-C2

Table III
Characteristics of the Sugar Ring Conformations and Glycosyl Torsion Angles of the dDMPs and dDMP-Na^+ obtained by PW91PW91/DZVP method[a].

Nucleoside sequence	dDMP-Na^+				dDMP			
	puckering	P	υ_{max}	χ	puckering	P	υ_{max}	χ
dCpdC	C2'-endo	168	31.2	-127	C2'-endo	166	30.5	-124
	C2'-endo	150	33.0	-135	C2'-endo	167	31.4	-127
dGpdG	C2'-endo	177	31.8	-109	C2'-endo	167	37.1	-124
	C2'-endo	173	31.3	-99	C3'-exo	206	30.1	-77
dCpdG	C2'-endo	167	30.6	-128	C2'-endo	167	31.3	-124
	C2'-endo	152	34.6	-103	C2'-endo	177	32.1	-93
dGpdC	C2'-endo	161	31.3	-122	C1'-exo	136	40.8	-139
	C2'-endo	153	34.0	-108	C3'-exo	185	30.3	-99
dG[b]	C3'-exo	190	28.4	-98				
dC[b]	C2'-endo	164	30.0	-128				

[a]P, pseudorotational phase angle of a sugar ring; υ_{max}, puckering amplitude; χ, glycoside torsion (O4'-C1'-N9-C4 for guanine and O4'-C1'-N1-C2 for cytosine containing nucleosides); the first row for each dDMP corresponds to nucleoside at the 5'-end; [b]experimental data, Ref. (22).

for cytosine containing nucleosides), which is in the range between 170° and 280°. Again, anionic dDMPs are characterized by a larger deviation of their glycosyl torsions in comparison with the data of the corresponding Na^+ complexes (see Table II). The glycosyl bond of the dDMPs deviates from the base ring plane by a few degrees, the exact value being dependent on the base sequence. The maximal deviations (up to 7°) happen in the structures characterized by inter-base hydrogen bonds, which will be discussed further.

An interesting feature of most of the obtained structures is a short contact (up to 3.3 Å) between O5′ and cytosine C6, which can be viewed as a weak C6-H6…O5′ hydrogen bond. A similarly short C…O distance was mentioned in the first X-ray diffraction resolved nucleoside, cytidine, as early as in 1950 (32). The crystal structure of the sodium salt of GpC (33) demonstrates similar short C6…O5′ distance of 3.29 Å. In some crystals this distance can be as short as 3.12 Å, *e.g.*, in 5′-O-(Guanosine-2′-O-phosphonomethyl)-cytidine tetradecahydrate (34). The obtained computational results can be considered as additional evidence of the importance of C-H…O weak H-bonds for nucleic acid structure. The computations clearly demonstrate that such H-bonds are not the result of crystal packing forces, but they are intrinsic features of these biomolecules.

The Geometry of Bases and Base-base Interactions

Good agreement between the computational and experimental data is obtained for the geometry of the bases. The majority of endo cyclic bond lengths and bond

Table IV
Endo cyclic Bond Angles of the Sugar Rings of the dDMPs and dDMP-Na^+ complexes.

Nucleoside sequence	dDMP-Na^+					dDMP				
	C1'	C2'	C3'	C4'	O4'	C1'	C2'	C3'	C4'	O4'
dCpdC	106.0	103.8	102.7	107.4	110.3	106.6	103.7	102.6	107.5	110.0
	105.1	103.3	103.0	107.7	109.6	106.0	103.4	103.1	106.9	110.5
dGpdG	106.2	103.4	102.8	106.5	110.8	106.0	102.0	101.6	107.0	109.0
	106.4	103.4	102.8	107.0	110.5	107.7	104.2	103.1	105.7	110.0
dCpdG	105.9	103.8	103.0	107.3	110.5	106.4	103.9	102.3	107.9	109.6
	105.1	102.5	102.9	107.4	109.8	106.7	102.8	102.7	106.5	110.6
dGpdC	105.7	103.8	103.0	107.3	110.3	104.5	102.1	102.8	107.5	106.3
	105.3	102.6	103.2	107.2	109.9	106.8	103.5	103.4	106.1	110.8
C2'-endo[a]	105.9	102.5	103.1	106.0	110.1					
C3'-endo[a]	106.8	102.4	102.2	104.5	110.3					

[a]Averaged crystal data [Ref. (22)] for the C2'-endo and C3'-endo conformations; the first row for each dDMP corresponds to the nucleoside at the 5'-end.

angles of the bases in the optimized structures differ from the averaged characteristics for monomer crystals (23) by 0.01-0.05 Å and 1-3°, respectively. Differences between the values of the same bond lengths in different structures are about or less than 0.01 Å, and those for valence angles differ by less than 1°. Similar results were obtained by using the B3LYP/6-31G(d,p) method. The base rings, which are normally expected to be planar, are slightly distorted according to our results, which can be seen in the endo cyclic torsions deviating from planarity by up to 5°. These effects are accompanied by significant pyramidalization of the amino groups with

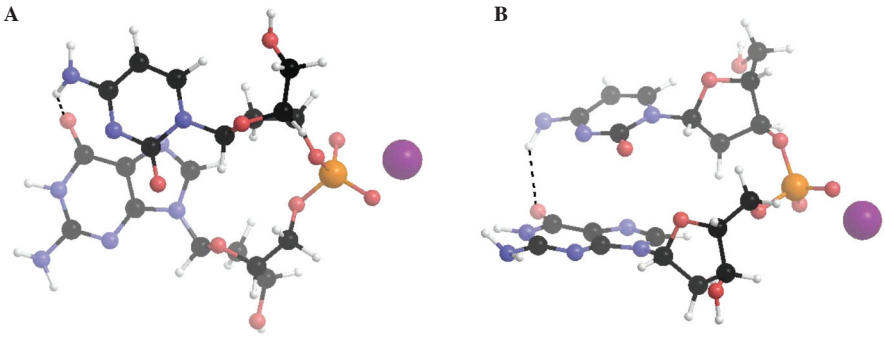

Figure 2: The minimum energy conformation of the dCpdG-Na^+ complex. Two orthogonal projections (panel **A** and panel **B**) illustrate the mutual disposition of the bases and inter-base H-bonding. Atoms closer to the observer are richer in color.

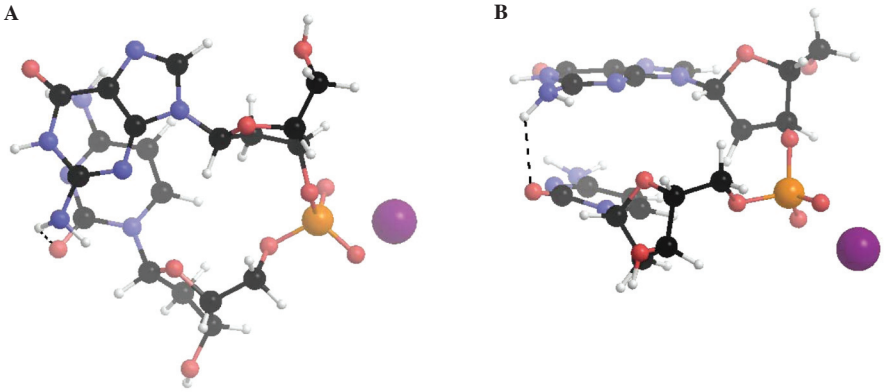

Figure 3: The minimum energy conformation of the dGpdC-Na^+ complex illustrating ring-on-ring stacking and N2-H21…O2 inter-base H-bonding.

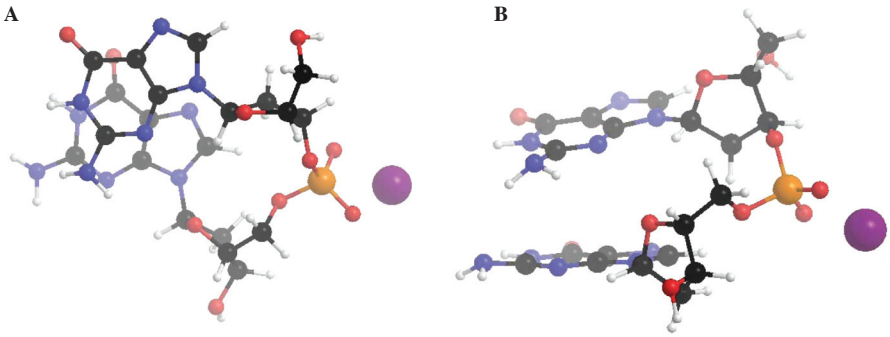

Figure 4: The minimum energy conformation of the dGpdG-Na^+ complex. The base-base ring-on-ring stacking resembles that of the purine-purine sequence in crystals of DNA duplexes.

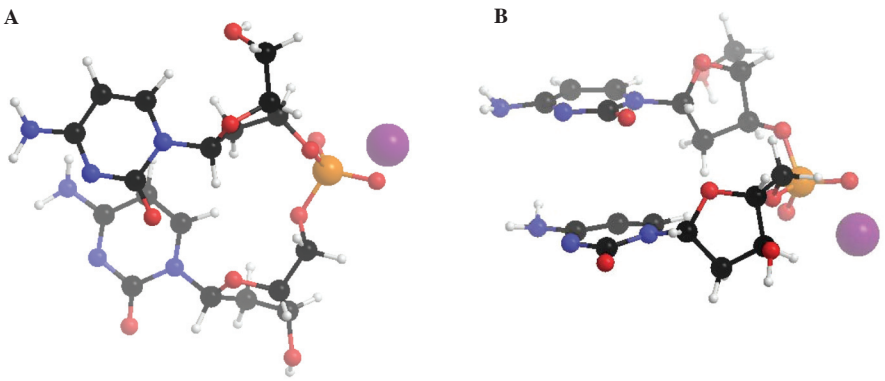

Figure 5: The minimum energy conformation of the dCpdC-Na^+ complex showing no base-base superposition.

particular details of the geometry depending on their interactions with other moieties of the dDMPs (see further discussion). It is well known that DFT methods in general tend to underestimate the level of pyramidalization of the amino group (24), because these effects are characterized by small energy differences. The present calculations revealing pyramidalization of the amino group lead us to believe that the energetic effect of the pyramidalization is sufficiently strong for it to be captured by DFT. Among the factors facilitating pyramidalization of the amino group is the novel H-bond type interaction observed between the stacked bases. As the calculations reveal, the dDMPs and their Na^+ complexes simultaneously containing Gua and Cyt in one molecule are characterized by formation of hydrogen bonds between these bases due to the pyramidalization of the amino groups. Such H-bonds can be observed between the amino group of the 5′-end base and the carbonyl oxygen of the 3′-end base, namely, in the dCpdG and the dCpdG-Na^+ (N4-H41…O6, Figure 2), and in the dGpdC and the dGpdC-Na^+ complexes (N2-H21…O2, Figure 3). Geometry characteristics of these H-bonds are listed in Table V. Remarkably, in the case of equivalent bases (dGpdG and dCpdC) the hydrogen atoms of the pyramidalized amino groups of both bases deviate in the same directions from their base planes (see Figures 4 and 5) leading to favorable dipole-dipole interactions of these groups.

Table V
Base-base hydrogen bonds in the dDMP and dDMP-Na+ complexes.

dDMP	H-bond X-H…Y	H…Y distance (Å)	X…Y distance (Å)	X-H…Y angle (°)
dGpdC	N2-H21…O2	2.26	3.10	139
dGpdCNa^+	N2-H21…O2	2.17	3.09	147
dCpdG	N4-H41…O6	2.18	3.05	141
dCpdG Na^+	N4-H41…O6	2.21	3.07	140

As seen from Figures 2-5 the base rings of the dDMPs are located in a nearly parallel disposition. The plane-plane angle between the stacked bases is in the range of 4.9 to 30.8 degrees. The largest deviation from ideal parallel orientation of the base rings is observed for the dGpdG anion (see Table I). This anion shows also the largest structural deviation of the conformational parameters from the experimental data, deviating considerably in this aspect from the other studied structures. Overall, the nearly parallel base – base arrangement takes place in all dDMP structures. According to the common wisdom such geometrical preference of the bases is usually attributed to favorable base-base dispersion interactions also known as “stacking interactions”. If such dispersion interactions are indeed important for the studied systems, then the level of the parallel disposition of the bases would be directly proportional to the degree of their spatial overlap. However, the analysis of the optimized structures of the bases points out that the parallel disposition of the bases is independent on the degree of the ring-on-ring overlap. Indeed, even in the least overlapping case of dCpdC anion and its Na^+ complex (see Figure 5) the base planes are ideally parallel to each other. Similarly, no significant influence is made on the parallel disposition of the bases by the presence or absence of the inter-base H-bonding. This suggests that the major contribution to the stability of the B-like single-stranded conformation comes from the sugar-phosphate backbone and only a minor contribution can be attributed to base-base stacking interactions.

Another characteristic feature of the dDMPs is related to the sequence of the bases. The dDMP anions with G at the 5′ end and their Na^+ complexes demonstrate extensive base ring-on-ring overlap (Figure 3 and 4), while the dDMP anions with C at the 5′ end and their Na^+ complexes have practically no base ring overlap (Figures 2 and 5). This observation is in excellent agreement with the Pur-Pur, Pur-Pyr, and Pyr-Pur arrangement in crystals of B-DNA duplexes (35). The difference in mutual base arrangements of the two dDMPs with the same base content, but different base sequence, *e.g.*, dGpdC and dCpdG, resembles that for the Pur-Pyr and Pyr-Pur steps of B-DNA duplexes in crystals (35). Namely, the first step demonstrates substantial ring-on-ring overlap (Figure 3). Base stacking can be seen in crystal structures of

sodium (33) and ammonium (36) salts of ribodinucleoside monophosphate GpC as well. At the same time there is no visible ring overlap in dCpdG where the exocyclic pyrimidine O2 is placed over the purine five-atom ring [Figure 2, see for comparison Figure 2 in Ref. (35)], presumably stabilizing the present geometry. A similar situation in the spatial disposition of the bases can be seen in dCpdG crystals (37). Considerable superposition of the two bases for dGpdG (Figure 4) and negligible inter-base overlap for dCpdC (Figure 5) are similar to the base-base arrangement taking place in the Pur-Pur and Pyr-Pyr steps of B-DNA duplexes (35).

Conclusions

The DFT geometry optimization studies of deoxydinucleoside monophosphate anions containing Gua and/or Cyt and their Na^+ complexes demonstrate existence of an important conformational minimum where the sugar ring puckering and sugar-phosphate backbone torsion angles resemble those of B-family DNA duplexes. All the sugar rings of the sodium complexes pertain to the C2′-endo type, while the sugars of the anions exhibit C1′-exo and C3′-exo puckerings as well. The heterocyclic rings of the bases retain a nearly planar configuration while the amino groups become non-planar, and their conformations depend on nucleoside content and sequence. The dGpdC, dCpdG dDMPs and their Na^+ complexes demonstrate formation of unusual hydrogen bonds between the carbonyl oxygen of one base and the pyramidalized amino group from the other base. The conformations obtained after geometry optimization show an almost parallel orientation of the bases with the interplane angle between the bases varying in the range from 4.9 to 30.8 degrees. It is found that the base-base superposition is independent on the extent of the inter-base overlap. From this it is concluded that the sugar-phosphate backbone contributes most to the stability of B-like single-stranded conformations and that base-base interactions play only a secondary role in the formation of these spatial structures. The dDMPs with the 5′ end G and their Na^+ complexes demonstrate extensive stacking overlap of the base rings, while in the dDMPs with the 5′ end C and their Na^+ complexes the inter-base overlap is practically absent. This observation correlates well with purine-purine, purine-pyrimidine, and pyrimidine-purine arrangements in crystals of B-DNA duplexes. The significant structural feature of the dDMPs is the formation of the cytosine base – phosphate oxygen (C6-H6…O5′) hydrogen bond, which is consistent with experimental observations.

References and Footnotes

1. A. D. MacKerell, Jr. *J Comp Chem 25*, 1584-1604 (2004).
2. V. I. Danilov, V. M. Anisimov, N. Kurita, and D. Hovorun. *Chem Phys Letters 412*, 285-293 (2005).
3. N. Kurita, V. I. Danilov, and V. M. Anisimov. *Chem Phys Letters 404*, 164-170 (2005).
4. M. Meyer, T. Steinke, M. Brandl, and J. Sühnel. *J Comp Chem 22*, 109-124 (2001).
5. J. Sponer, J. Leszczynski, and P. Hobza. *Biopolymers 61*, 3-31 (2002).
6. O. V. Shishkin, L. Gorb, O. A. Zhikol, and J. Leszczynski. *J. Biomol. Struct. Dyn. 22*, 227-244 (2004).
7. O. V. Shishkin, G. V. Palamarchuk, L. Gorb, and J. Leszczynski. *J Phys Chem B 110*, 4413-4422 (2006).
8. V. I. Poltev, V. M. Anisimov, V. I. Danilov, A. Deriabina, E. Gonzalez, A. Jurkiewicz, A. Les, and N. Polteva. *J Biomol Struct Dyn 24*, 660 (2007).
9. C. Sosa, J. Andzelm, B. C. Elkin, E. Wimmer, K. D. Dobbs, and D. A. Dixon. *J Phys Chem 96*, 6630-6636 (1992).
10. N. Godbout, D. R. Salahub, J. Andzelm, and E. Wimmer. *Can J Chem 70*, 560-571 (1992).
11. P. C. Hariharan and J. A. Pople. *Theor Chim Acta (Berlin) 28*, 213-222 (1973).
12. J. P. Perdew and Y. Wang. *Phys Rev B 45*, 13244-13249 (1992).
13. N. Kurita, H. Inoue, H. Sekino. *Chem Phys Letters 370*, 161-169 (2003).
14. A. D. Becke. *J Chem Phys 84*, 4524-4529 (1986).
15. M. Swart, T. van der Wijst, C. F. Guerra, and F. M. Bickelhaupt. *J Mol Model 13*, 1245-1257 (2007).
16. A. D. Becke. *Phys Rev A 38*, 3098-3100 (1988).
17. A. D. Becke. *J Chem Phys 98*, 5648-5652 (1993).
18. C. Lee, W. Yang, and R. G. Parr. *Phys Rev B 37*, 785-789 (1988).
19. M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S.

Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, Jr. *J Comp Chem 14*, 1347-1363 (1993).
20. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. Montgomery, J. A., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford CT (2004).
21. G. Sun, J. H. Voigt, I. V. Filippov, V. E. Marquez, and M. C. Nicklaus. *J Chem Inf Model 44*, 1752-1762 (2004).
22. A. Gelbin, B. Schneider, L. Clowney, S. H. Hsieh, W. K. Olson, and H. M. Berman. *J Am Chem Soc 118*, 519-529 (1996).
23. L. Clowney, S. C. Jain, A. R. Srinivasan, J. Westbrook, W. K. Olson, and H. M. Berman. *J Am Chem Soc 118*, 509-518 (1996).
24. W. Koch and M. C. Holthausen. *A Chemist's Guide to Density Functional Theory, 2nd Ed*. Wiley (2001).
25. C. Møller and M. S. Plesset. *Phys Rev 46*, 618-622 (1934).
26. M. Head-Gordon, J. A. Pople, and M. J. Frisch. *Chem Phys Letters 153*, 503-506 (1988).
27. T. Clark, J. Chandrasekhar, G. W. Spitznagel, and P. V. R. Schleyer. *J Comp Chem 4*, 294-301 (1983).
28. T. H. J. Dunning. *J Chem Phys 90*, 1007-1023 (1989).
29. R. A. Kendall, T. H. Dunning, Jr., and R. J. Harrison. *J Chem Phys 96*, 6796-6806 (1992).
30. A. F. Jalbout and L. Adamowicz. *Chem Phys Letters 420*, 209-214 (2006).
31. C. Altona and M. Sundaralingam. *J Am Chem Soc 94*, 8205-8212 (1972).
32. S. Fuberg. *Acta Cryst 3*, 325-333 (1950).
33. J. M. Rosenberg, N. C. Seeman, R. O. Day, and A. Rich. *J Mol Biol 104*, 145-167 (1976).
34. J. Zachova, I. Cisarova, M. Budesinsk, R. Liboska, Z. Tocik, and I. Rosenberg. *Nucleosides, Nucleotides and Nucleic Acids 18*, 2581-2599 (1999).
35. R. E. Dickerson. In *Structure, Motion, Interaction and Expression of Biological Macromolecules*, Vol. 1, pp. 17-36. Eds., R. H. Sarma and M. H. Sarma. Adenine Press, Albany, NY (1998).
36. A. Aggarwal, S. A. Islam, R. Kuroda, M. R. Sanderson, S. Neidle, and H. M. Berman. *Acta Cryst B 39*, 98-104 (1983).
37. B. Ramakrishnan and M. A. Viswamitra. *J Biomol Struct Dyn 6*, 511-523 (1988).

Date Received: July 26, 2007

Communicated by the Editor Maxim Frank-Kamenetskii

Journal of Molecular Structure: THEOCHEM 912 (2009) 53–59

Contents lists available at ScienceDirect

Journal of Molecular Structure: THEOCHEM

journal homepage: www.elsevier.com/locate/theochem

DFT study of minimal fragments of nucleic acid single chain for explication of sequence dependence of DNA duplex conformation

Valery I. Poltev [a,b,*], Victor M. Anisimov [c], Victor I. Danilov [d], Alexandra Deriabina [a], Eduardo Gonzalez [a], Dolores Garcia [a], Francisco Rivas [a], Agata Jurkiewicz [e], Andrzej Leś [e], Nina Polteva [b]

[a] Universidad Autonoma de Puebla, Av. San Claudio y Rio Verde, Puebla 72570, Mexico
[b] Institute of Theoretical and Experimental Biophysics RAS, Pushchino 142290, Russia
[c] University of Maryland, 20 Penn St., Baltimore, MD 21201, USA
[d] Institute of Molecular Biology and Genetics, Kyiv 03143, Ukraine
[e] University of Warsaw, Warsaw 02-093, Poland

ARTICLE INFO

Article history:
Received 12 November 2008
Received in revised form 19 March 2009
Accepted 24 March 2009
Available online 29 March 2009

Keywords:
DFT
dDMP
Conformation
DNA

ABSTRACT

DFT studies of desoxydinucleoside monophosphate complexes with Na-ions (dDMPs), using DGAUSS and ADF software reveal that the characteristic B-type conformation of DNA duplexes is initially predisposed in single strand in the form of a local minimum. Such predisposition may be important for biological functioning of DNA by facilitating an addition of new nucleotide in the processes of DNA duplication and restoration of double helix after unwinding. The computational results demonstrate the important contribution of intra-strand interactions to conformational stability and sequence-dependent variability of B-DNA duplexes. The main characteristics of the "canonical" BI conformations, such as the regions of torsion angles of sugar-phosphate backbone, C2′-endo or very close to C2′-endo sugar puckerings, and nearly parallel base-ring arrangements are already present in the single strand and are universal for all 16 dDMPs. The dihedral angle between the bases does not depend on the extent of their overlap, which suggests the major contribution of sugar-phosphate backbone to formation of these minima. The study reveals characteristic sequence dependence of the base arrangements. Namely, an extensive base-ring overlap takes place for all purine–purine and purine–pyrimidine sequences, and negligible ring overlap is observed in all pyrimidine–purine and pyrimidine–pyrimidine sequences. These arrangements closely resemble the geometrical characteristics of BI DNA in duplex crystals. Preliminary results of search for other energy minima related to helical DNA structures revealed the existence of BII-like minima for some dDMPs.

1. Introduction

Study of the role of nucleic acids building blocks in conformational stability and conformational preferences of DNA with specific base sequences is an important step toward understanding the molecular mechanisms of genetic processes. Experimental studies revealed various patterns of three-dimensional structure of DNA, including well-known A-, B-, and Z- double helices, which differ in sugar-phosphate torsional angles and mutual base arrangements as presented in the Nucleic Acid Data Bank [1]). The conformational preferences of DNA duplex fragments can be studied by experimental, e.g. X-ray and NMR, and theoretical methods, including molecular mechanics and quantum mechanics computations. Conventionally, the conformational stability of the double helix is attributed to inter-strand hydrogen-bond interactions of complimentary bases and to inter- and intra-strand stacking interactions. Interactions of the duplex with surrounding media, including water molecules and counterions, further modulate the duplex conformation inside the region of its possible geometric variations. The conformational preferences of nucleotide sequences have been derived from X-ray diffraction studies of DNA duplex fragments containing from 6 to 14 nucleotide pairs, followed by theoretical considerations and computer simulations. Hundreds such studies are collected in Nucleic Acid Data Bank [1] and analyzed in a recent paper [2].

Contrary to duplexes the spectrum of conformational possibilities of single chain fragments is more difficult to address by experimental methods. The question, in what degree the single-strand conformational preferences determine the duplex conformation, are hardly possible to address by experimental studies. To large extent the conformation of single-strand fragments is determined by

* Corresponding author. Address: Universidad Autonoma de Puebla, Av. San Claudio y Rio Verde, Puebla 72570, Mexico. Tel.: +52 2225765578.
E-mail addresses: poltev@fcfm.buap.mx, vpoltev@yahoo.com (V.I. Poltev).

0166-1280/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.theochem.2009.03.022

their interactions with other components of the crystal or solution, thus, one need to involve first principle theoretical methods to derive unperturbed conformational characteristics of the single-strand chains. To evaluate the contribution of intra-strand interactions to variation of duplex structures we performed quantum mechanical (QM) calculations of minimal fragments of DNA chains using density functional theory (DFT) methods. The latter provide a reasonable compromise between accuracy and resource demand among available computational methods. In general, *ab initio* methods, e.g. MP2 and CCSD provide more accurate results but they are computationally too expensive to be used for such large systems.

Deoxydinucleoside monophosphate is the minimal fragment of DNA single strand which consists of two nucleotides situated along the chain. Correspondingly, dDMP determines the conformation of the single strand via inter-strand interaction of the nucleotides. Due to the computational cost, larger units and particularly the systems of practical interest can only be approached using molecular mechanics. The success of such computation vitally depends on refinement of the force fields and on their ability to quantitatively reproduce the detailed conformational aspects of the minimal fragments. The need for such refinement of molecular mechanics force fields has recently become apparent from computations of DNA subunits and duplex fragments (see e.g. long molecular dynamics simulations of tetranucleotide repeating sequences [3,4]). These "simple" fragments have extremely complicated conformational profile due to the existence of many degrees of freedom, including six sugar-phosphate torsions, glycoside torsions, and highly flexible deoxyribose ring. As it was mentioned by Svozil et al. [5] quantum-mechanical computations of a simple fragment of sugar-phosphate backbone having no connected bases is already not trivial because of complexity of its conformational space.

In the light of these observations we restricted ourselves to consideration of B-family like structures of deoxydinucleoside monophosphates applying special emphasis on dependence of their three-dimensional structure on nucleotide sequence.

In our previous work [6] we considered four deoxydinucleoside monophosphates and their sodium-complexes containing guanine and cytosine bases. We obtained energy minima corresponding to "canonical" B conformations of DNA duplexes and recognized the principal contribution of sugar-phosphate backbone to parallel arrangement of bases, which is sequence dependent. The neutralized deoxydinucleoside monophosphates–Na^+ complexes (dDMPs) found to be better models of the minimal DNA fragments in crystals and solutions in comparison to their anionic counterparts, as their calculated geometries were closer to experimental data [6] than those of anionic systems. In this study we extended our previous calculations to include all possible 16 dDMPs. Using different computational methods, we confirmed the general conclusions derived in our previous study and found new regularities in the dependence of dDMPs configuration on the base sequence. Among these findings is the observation of another conformational energy minimum for B-family duplexes, namely BII conformation. It has some sugar-phosphate torsions located in regions different from those for the canonical BI conformation [2].

2. Methods

The systematic study of all naturally occurring 16 dDMP–Na^+ complexes was performed using the DGAUSS program [7]. We considered this program due to its remarkable performance and acceptable accuracy as described earlier [6]. DGAUSS uses the so-called fitting basis set for fast approximation of the Coulomb energy, facilitating QM study of relatively large systems. We used split-valence double zeta basis set augmented with polarization function (DZVP), which is specifically optimized for DFT calculations [7,8]. The original gradient-corrected exchange-correlation functional developed by Perdew and Wang PW91PW91 [9] is used in this work. This functional is among few which describe van der Waals interactions in dimers of rare gases [10,11] and hydrocarbons [10], as well as in hydrogen-bonded complexes of simple molecules including water–water and water–formamide dimers [10]. Tsuzuki and Lüthi came to a conclusion that despite of the difficulties the PW91 functional has with dispersion interactions, it may be a viable alternative to the *ab initio* methods [10]. More recent studies demonstrate that PW91 functional reasonably reproduces correlated *ab initio* results for hydrogen-bonded [12,13] and stacking [14] complexes of DNA bases. In addition, the results of our study (see below) demonstrate that the conformational characteristics of the considered systems are governed mainly by properties of the sugar-phosphate backbone rather than by base stacking. Good agreement with *ab initio* MP2 results obtained for separate bases (see Section 3.3) is an additional argument in favor of the selected functional.

In our previous work [6], to test the sensitivity of the obtained results to the choice of a particular DFT functional, basis sets, and density fitting, additional minimization of selected structures was performed using the 6-31G(d,p) basis set [15], and the B3LYP functional [16] as implemented in the GAMESS program [17] and in the GAUSSIAN 03 program package [18]. The resulting structures had their conformational parameters only slightly deviating from the initial ones contorting to a few degrees in torsion and valence angles, and no more than 0.02 Å in bond lengths, thereby supporting the main conclusions made in that work. Similar to that the conclusions of the present work were validated by using ADF (Amsterdam Density Functional) program [19] and triple zeta basis set. Basically the agreement with experimental data only improves upon improvement in the DFT functional and basis set. The ADF package [19] uses Slater functions and accurate and tunable numerical integration scheme, in which the grid is automatically, adapted to the available basis functions. A density fit and linear scaling techniques reduce the cost of the Coulomb potential evaluation. The use of this software with PW91 functional and TZP (triple zeta with polarization functions) basis set enabled us to reproduce experimental phosphate group geometry somewhat better than that in our previous computations; however the main conformational features of dDMPs obtained via DGAUSS program were conserved.

Starting BI-like geometries of dDMPs for the first optimizations of each of the dDMPs were generated using the HyperChem program version 7.01 (HyperCube, Inc.). Due to the limitations in this version of the HyperChem software the sugar rings were constructed with C3′-exo instead of C2′-endo puckering, characterized by a pseudorotation phase angle $P = 192^\circ$. In our previous work [6] we ruled out the possibility of biasing the computational results toward the initial sugar puckering form. In the present work we reoptimized the structures of dDMPs containing Guanine or Cytosine bases starting from their last optimized geometries. In addition to that, for some of the dDMPs, the optimizations were started from experimental geometries obtained from Nucleic Acid Data Bank [1]. The initial BII-like configurations were obtained from the experimental duplex fragments with BII characteristic sets of torsion angles. Sodium cations of the dDMPs were placed at equidistant positions to the anionic oxygens of phosphate group (at O...Na distance of 2.2 Å) in the O1P–P–O2P plane. The generated structures were subjected to all-atom geometry optimization and the geometric characteristics of the optimized minimum energy conformations including bond lengths, valence and torsion angles as well as hydrogen-bond characteristics (for the structures with H bonds) and distances from Na^+ ion to anion oxygens were

analyzed. The program PROSIT [20] was used for the analysis of sugar-ring puckering and glycosidic torsions.

3. Results and discussion

3.1. General features of B-like minimum energy conformations. Geometry of sugar-phosphate backbone

Following our previous work the present calculations confirm our supposition that for all dDMPs there exist energy minima corresponding to B-family conformations of DNA duplexes. The torsion angles of sugar-phosphate backbone, sugar-ring puckering, and mutual positions of the bases of these minima are characteristic for BI-conformations of DNA fragments in crystals. The bond lengths and valence angles of all the optimized structures are within the limits of the experimentally observed values [21,22] with the exception of some values for the phosphate group. The latter differences can be partially attributed to the differences in surrounding of anionic groups of the model dDMPs relative to the crystals or solutions as discussed in our previous paper [6].

The calculated values of sugar-phosphate torsions of the BI minima for all the 16 dDMPs are listed in Table 1 along with experimental mean values for dinucleoside monophosphate (DMP), and trinucleoside diphosphate (TMP) crystals [22], and averaged values for BI-conformation steps of olygonucleotide duplexes [2]. For some of the sequences characterized by the most pronounced deviation of one of conformational characteristics from other sequences, we performed additional optimizations starting from several different initial geometries generated by displacing some of the torsions from the values of previously optimized structures. These displacements were limited in magnitude to a few degrees for the structures to remain within the BI-conformation region. The differences found between the previous conformations and the corresponding new optimized structures were at the range of 0.1–0.4 kcal/mol for total energy and 1–4° for torsions. It points to the existence of a wide region of BI conformations in the multi-dimensional space of conformational variables. The conformational characteristics listed in the paper correspond to the lowest energy values.

The use of ADF software with the PW91 functional and TZP basis set minimally changed the values of torsion angles. Some of these changes are less than 1° with the maximal deviation up to 9°; the averaged values for 7 torsions mentioned in Table 1 vary from 1 to 4° depending on the base sequence. Despite the individual differences all the torsions are preserved in the regions characteristic for BI family of duplex conformations.

To simplify the analysis of the results and comparison between different sequences, we subdivided the 16 dDMPs to four groups corresponding to base sequences, purine–purine (PUR–PUR), pyrimidine–pyrimidine (PYR–PYR), purine–pyrimidine (PUR–PYR), and pyrimidine–purine (PYR–PUR) sequences. Each of the analyzed torsion angles stayed in the limits of one of the regions observed for DMPs and TDPs in crystals [22], close to averaged BI-values [2], namely, *−sc* (*−gauche*) for α and ξ, *ap* (*trans*) for β and ε, *+sc* (*+gauche*) for γ; δ values are characteristic for C2′-endo sugar puckering. Variation in α dihedral angles is very small along the series of base sequences, being within 3.9° from the experimental value; γ dihedrals vary from 53.1 to 58°, and $\delta 1$ torsions vary in 8.2° interval (Table 1). It is remarkable to see very small variations in some torsions within one dDMP group, e.g. γ values are preserved within 1.5° for PUR–PUR and 1.2° for PYR–PYR sequences; ε values vary in 1.3° interval for PYR–PYR and PUR–PYR sequences. One may see remarkable systematic changes in some torsions upon going from one dDMP group of base sequences to another, e.g. all the values of $\delta 1$ angle for PUR–PUR sequences are greater than those for PYR–PYR group, and the values of $\delta 1$ torsions for all the PYR–PYR sequences are greater than those for PUR–PYR or PYR–PUR groups. Notable are the ξ dihedral values in PYR–PUR sequences, which are somewhat greater than those for other dDMPs. The systematic variations of torsion angles with base sequence result in specific patterns of mutual base arrangements, which will be discussed below.

Table 1
Torsion angles of the sugar-phosphate backbone (definitions are in Fig. 1) and average values of the angles between the base planes (φ) for the calculated BI conformations of dDMP–Na^+ complexes.

dDMPs	$\delta 1$	ε	ζ	α	β	γ	$\delta 2$	φ
Purine–Purine								
dApdA	141.7	175.4	263.5	295.3	191.9	54.9	137.9	3.1
dGpdG	141.7	173.9	259.2	298.0	190.7	54.5	141.8	14.1
dAdpG	142.5	178.0	261.6	297.4	189.6	52.3	137.9	6.1
dGdpA	143.8	178.5	263.3	295.8	191.6	52.7	140.6	8.0
Pyrimidine–Pyrimidine								
dCpdC	139.1	180.7	262.2	294.3	185.8	57.9	133.6	4.8
dTpdC	139.7	181.5	262.1	294.2	184.7	58.5	132.9	3.6
dCpdT	138.9	182.0	263.3	295.9	184.4	58.0	130.2	11.8
dTpdT	139.1	181.5	264.2	295.4	185.5	56.6	129.9	11.2
Purine–Pyrimidine								
dApdC	136.5	183.1	259.8	297.5	181.2	56.8	132.2	18.1
dApdT	137.1	181.9	263.7	296.1	181.6	57.5	129.6	8.2
dGpdT	138.4	182.3	262.9	297.0	181.5	55.2	130.3	8.7
dGpdC	136.8	182.0	260.6	298.8	183.6	55.3	134.6	11.9
Pyrimidine–Purine								
dCpdA	136.1	178.9	267.9	295.6	183.9	53.9	134.4	22.8
dTpdG	135.8	182.4	266.5	296.3	181.1	54.1	131.8	18.7
dCpdG	138.1	179.7	265.2	296.9	184.7	53.1	134.9	21.2
dTpdA	135.2	182.4	265.7	297.4	179.5	55.5	129.7	20.2
Exp.[a]	145.2	214.0	289.2	285.3	183.5	52.5	145.2	
Exp. SD[a]	4.9	8.6	4.8	9.8	13.0	5.7	4.9	
BI[b]	132.8	181.7	263.2	299.0	179.3	48.4	132.8	

[a] Experimental data from Ref. [22].
[b] Averaged values for BI conformations from Ref. [2].

3.2. Sugar-ring puckering and sugar-base mutual orientation

Sugar ring bond lengths and valence angles for all the calculated dDMP structures are close to those in the averaged crystal data [21,22]. DGAUSS results on sugar-ring puckering of BI conformations are presented in Table 2 along with their numerical characteristics in terms of pseudorotation concept of Altona and Sundaralingam [23]. Computational results on pseudorotation phase angle *P*, obtained using ADF program package differ from DGAUSS results collected in Table 2 by less than 10°. The sugar-ring puckering of all the deoxyriboses pertain to C2′-endo type or very close to it. Rather small but systematic differences between 5′-end and 3′-end nucleoside sugars can be seen from Table 2. The pseudorotation phase angle *P* for the first sugar varies from 161° to 181°, while for the second one this value varies from 141° to 172°. One of 16 dDMPs displays C3′-exo puckering of C5′-end sugar with *P* values close to those for C2′-endo region, and 4 dDMPs have C1′-exo conformations of the 3′-end sugars with *P* value being few degrees smaller than those of C2′-endo sugar conformations. For the dDMPs with sugar puckering close to C2′-endo border, we performed additional calculations starting from the structures generated with C2′-endo sugars to confirm that the small deviations of sugar puckering are real features rather than a results of limited energy minimization. For all the dDMPs the *P* value of 5′-end sugar is greater than that for 3′-end nucleoside. Puckering amplitudes fall in the most populated region for monomer crystals, namely from 28.6° to 37.3°.

The mutual orientations of the sugar and base rings correspond to the anti region of the glycosyl torsion angle χ (O4′–C1′–N9–C4

Table 2
Characteristics of the sugar ring conformations and glycosyl torsion angles of the dDMP–Na^+ complexes obtained by PW91PW91/DZVP method.

dDMP–Na^+	5′-end nucleoside				3′-end nucleoside			
	Puckering	P	υ_{max}	χ	Puckering	P	υ_{max}	χ
Purine–Purine								
dApdA	C2′-endo	176	31.7	−115	C2′-endo	167	30.3	−104
dGpdG	C2′-endo	177	31.8	−109	C2′-endo	173	31.3	−99
dApdG	C2′-endo	177	31.9	−116	C2′-endo	168	30.1	−105
dGpdA	C3′-exo	181	31.6	−116	C2′-endo	172	31.6	−100.0
Pyrimidine–Pyrimidine								
dCpdC	C2′-endo	168	31.2	−127	C2′-endo	150	33.0	−135
dTpdC	C2′-endo	168	31.5	−126	C2′-endo	150	34.0	−133
dCpdT	C2′-endo	167	31.2	−133	C1′-exo	143	37.1	−126
dTpdT	C2′-endo	169	30.9	−127	C1′-exo	142	37.2	−126
Purine–Pyrimidine								
dApdC	C2′-endo	160	31.6	−121	C2′-endo	148	34.1	−110
dApdT	C2′-endo	164	30.8	−120	C1′-exo	143	37.9	−109
dGpdT	C2′-endo	169	29.9	−120	C2′-endo	145	34.4	−112
dGpdC	C2′-endo	161	31.3	−122	C2′-endo	153	34.0	−108
Pyrimidine–Purine								
dCpdA	C2′-endo	166	28.6	−129	C2′-endo	148	36.9	−105
dTpdG	C2′-endo	167	28.2	−128	C2′-endo	145	35.1	−109
dCpdG	C2′-endo	168	30.6	−128	C2′-endo	152	34.6	−103
dTpdA	C2′-endo	164	28.7	−130	C1′-exo	141	36.6	−108

for purine and O4′–C1′–N1–C2 for pyrimidine containing nucleosides). This angle varies in the range between 225° and 252° for pyrimidines in agreement with experimental finding of 229.8 ± 18.4°; for purines this angle varies between 238° and 261° in agreement with experimental value of 237 ± 24.3° [22]. The angle χ is more variable in 3′-end nucleoside as compared to 5′-end one. The glycosyl bonds of the most of the nucleosides deviate slightly from the base-ring plane, and for some dDMP the deviation can be up to 8°, the exact value being dependent on the base sequence.

In the previous paper [6] we mentioned a short contact (up to 3.3 Å) between O5′ and cytosine C6, which can be viewed as a weak C6–H6...O5′ hydrogen-bond. A similarly short C...O distance was mentioned in the first X-ray diffraction resolved nucleoside, cytidine, as early as in 1950 [24]. Analysis of the more representative set of optimized structures confirms the presence of the short C6...O5′ contact in pyrimidine nucleosides and similar C8...O5′ contact in purine ones. Our computations clearly demonstrate that such weak H-bond is an intrinsic feature of the nucleic acid subunits. Naturally, because these distances correspond to 5′-end nucleoside they cannot be considered as a structural characteristic of inter-nucleoside region on the polynucleotide chain.

3.3. Geometry of the bases and the patterns of mutual base arrangements

Good agreement between computational and experimental data is obtained for the geometry of purine and pyrimidine rings. The majority of endo cyclic bond lengths and bond angles in the optimized structures differ from the averaged characteristics for monomer crystals [21] by 0.004–0.015 Å and 1–3°, respectively. Even closer agreement is observed when our DFT calculations are compared against the more accurate MP2 *ab initio* calculations of bases [25]. Differences between the values of bond lengths in different structures are about or less than 0.01 Å, and those for valence angles are less than 1°. According to our calculations the base rings of all the dDMPs are slightly distorted from planarity, which can be seen in deviation of some endo cyclic torsions from expected values for planar structure by up to 6°. The extent of nonplanarity of the base rings varies from one nucleotide sequence to another. The purine and pyrimidine rings in the optimized structures of the separate bases are practically planar with all the intra-ring torsions being within the ±1° limit.

A significant pyramidalization of the amino groups is observed in all the optimized structures despite the general tendency of DFT methods to underestimate the level of pyramidalization [26]. The degree of pyramidalization of the amino groups depends on base sequence in dDMP. It is interesting to note that our calculations of separate bases reproduce well the geometry characteristics of pyramidalized amino groups obtained by so far most accurate calculations performed at the MP2 level with a large basis set. These characteristics demonstrate substantially more pronounced nonplanarity of guanine amino group in comparison with adenine and cytosine ones. The sum of three valence angles of amino group calculated by DGAUSS program equals to 352.3, 354.0, and 339.9° for cytosine, adenine, and guanine, respectively. These values calculated by *ab initio* MP2/aug-cc-pVQZ method are 355.4, 355.2, and 342.0° [25], and those obtained from *ab initio* MP2/6-31G(2df,p) level of calculations are 351.9, 352.9, and 339.6°, respectively [27]. The close agreement with MP2 results was obtained for torsion angles of amino group. For example we can mention the torsion angle describing out of plane displacement of one of amino group hydrogens, namely (N of the ring)–(C of the ring)–(amino N)–(H participating in Watson–Crick pairing). The values calculated by DGAUSS program for Cyt, Ade, and Gua are 14.5, 13.1, and 39.7°, respectively. They agree well with the values of 10.0, 12.9, and 36.5° obtained via MP2/aug-cc-pVQZ method [25], as well as with 12.6, 16.5, and 39.2° obtained from MP2/6-31G(2df,p) calculations [27]. All the dDMPs have more pronounced amino group pyramidalization as compared to that of the separate bases. The sum of three valence angles of amino group is confined to the region from 332.9 to 347.9, from 336.7 to 354.6, and from 331.0 to 339.8° for Cyt, Ade, and Gua bases of dDMPs respectively. The torsion angle mentioned above varies from 13.3 to 24.3, from 16.1 to 23.1, and from 14.7 to 51.1 for Cyt, Ade, and Gua, respectively. Such pyramidalization of amino group enables the formation of N–H...O hydrogen-bonds between the amino group of the 5′-end base and the carbonyl oxygen of the 3′-end base.

Two base rings in optimized structures of all the dDMPs are located in nearly parallel disposition, and the plane-plane angle between the rings of the stacked bases varies from 3° to 23° (see

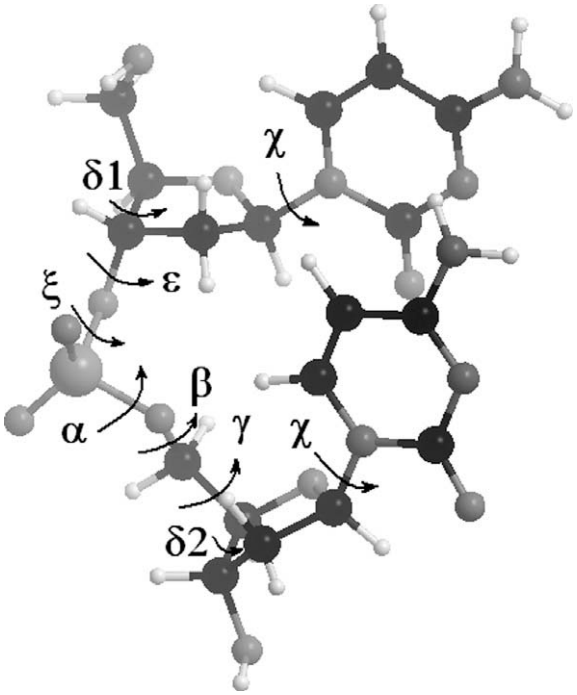

Fig. 1. Designation of the torsion angles of the sugar-phosphate backbone of the dDMPs.

Table 1). Such base–base arrangement is usually attributed to dispersion interactions between the planar purine and pyrimidine rings. Contrarily to that our results demonstrate, that these interactions do not play the crucial role in the case of the dDMPs. Indeed, as follows from the calculations, the angle between base planes does not depend on the base overlap; nearly parallel plane arrangement occurs for the structures, which have practically no base overlap (dTpdC and dCpdC) and also for the structures with extensive base overlap (dApdA and dGpdA). From this we conclude that the existence of BI-like local energy minima is not determined by the base–base stacking interactions but the major contribution to the stability of the B-like single-stranded fragments comes from the conformational properties of the sugar-phosphate backbone.

The analysis of mutual base positions in the optimized structures of four dDMPs containing Gua and Cyt enabled us to previously conclude that base overlap depends on nucleotide sequence, namely, extensive overlap in Gua-Gua and Gua-Cyt sequences but minor overlap in Cyt-Cyt and Cyt-Gua sequences [6]. This observation is in excellent agreement with the base disposition patterns extracted from crystallographic structures of B-DNA duplexes [28]. Based on these data we conclude that the mutual position of the bases in minimal fragments of single DNA strands predisposes the PUR–PUR, PUR–PYR, PYR–PUR, and PYR–PYR base arrangements in DNA duplexes. The calculations for all the 16 dDMPs demonstrate that PUR–PUR and PUR–PYR base sequences have extended overlap of pyrimidine rings of the bases. An example of such structure, the optimized dApdT conformation in comparison with dApdT step of the d(CpGpCpGpApApTpTpCpGpCpG) duplex [29] is displayed in Fig. 2. Our optimization of this dDMP was initiated from the standard sequence-independent B conformation fragment. However during the calculation the studied structure closely converged to the experimental structure pattern of PUR–PYR sequence of the above DNA fragment (see also Fig. 2d of [28] demonstrating the dApdT fragment from another X-ray structure [30]). Interestingly all dDMPs with PYR–PUR and PYR–PYR base sequence have small to almost negligible overlap of the bases. This can be seen on the example of dCpdA, with the optimization started from the sequence-independent structure. The resulting structure is displayed in Fig. 3 in comparison with the corresponding step of the d(CpGpCpApApTpTpGpCpG) duplex [31] (see also Fig. 2a of [28] demonstrating the dCpdA fragment from another X-ray structure [32]). This result enables us to extrapolate the important role of local interactions in the single chain going to the global sequence dependence of double helix, or more specifically, one can speak about the conformational preferences of sugar-phosphate backbone of single chain contributing to the conformational preferences of the fragments of DNA duplexes. The existence of such minima for single chain fragments may be important for biological functioning of DNA facilitating an addition of new nucleotide in the processes of its replication and restoration of double helix after unwinding.

3.4. Perspectives of the investigation. Other minima of DMPs related to the duplex structures, BII-like conformations

The DFT calculations for all the 16 dDMPs have confirmed our supposition about existence of B-like local energy minimum in single-chain DNA fragment. Surely, the use of more accurate QM

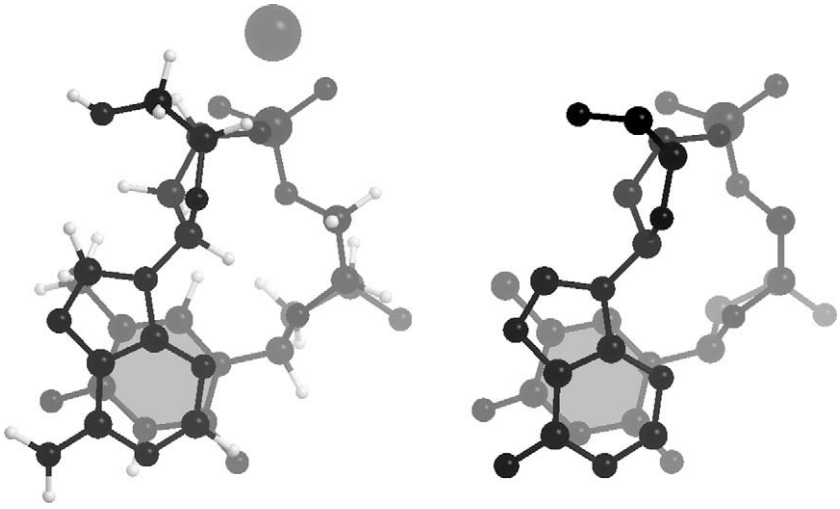

Fig. 2. The minimum energy conformation of the dApdT–Na^+ complex (left), and dApdT step of d(CpGpCpGpApApTpTpCpGpCpG) duplex in the crystal [22] (right). Extensive base–base overlap and nearly the same mutual base positions can be observed. Atoms closer to the observer are richer in color.

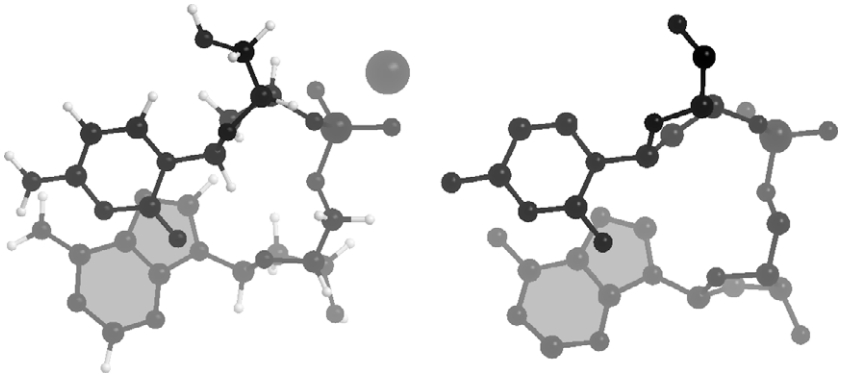

Fig. 3. The minimum energy conformation of the dCpdA–Na^+ complex (left), and dCpdA step of d(CpGpCpApApTpTpGpCpG) duplex in the crystal [24] (right).

Table 3
Torsion angles of the sugar-phosphate backbone and average values of the angles between the base planes (φ) for the calculated BII conformations of dDMP–Na^+ complexes.

dDMP–Na^+	$\delta 1$	ε	ζ	α	β	γ	$\delta 2$	φ
dCpdA	130.7	228.4	194.6	300.5	155.1	49.3	142.5	10.6
dCpdG	128.8	242.8	171.3	295.8	136.7	50.9	139.0	6.45
dGpdG	138.7	224.2	193.1	299.5	146.8	47.9	129.7	4.9
BII[a]	143.0	251.1	168.0	292.6	143.1	46.0	143.0	

[a] Averaged values for BII conformations from Ref. [2].

methods and the consideration of larger fragments could result in uncovering additional structural details leading to discovery of new regularities determining the stability of three-dimensional structure of DNA. But the main conclusions about the conformational predisposition of single-chain strands to B-type DNA helix formation are unlikely to be changed. Other questions arise in relation to these calculations and DNA structure. The first set of the problems can be formulated as follows: "Is this situation with single-chain interactions is unique to the canonical B form?" What about other DNA conformations revealed in the crystals of duplex fragments, e.g. A form, BII conformations, and mixed A/B or B/A forms of dinucleotide steps? Is it possible to predict from computations of dDMPs the preferences of various steps to the specific conformation or the contributions of other interactions become predominant? To answer these questions we started to search for other minima of dDMPs related to helical structures. Such a study is an order of magnitude more extensive and time-consuming one as compared to those presented here. Nevertheless, some preliminary results and further suggestions can already be mentioned. As for A form, preliminary calculations confirm our expectation of such local minima. These minima are deeper than those for B form (e.g. for dCpdC A form is more favorable than B form by about 5 kcal/mol). We detected the minima corresponding to BII conformation for three dDMPs, namely to dCpdG, dCpdA, and dGpdG, starting the optimization from the geometry of these steps directly extracted from the X-ray structures. The torsion angles of sugar-phosphate backbone and inter-base angles are listed in the Table 3 together with averaged values of torsions for BII structure [2]. These results definitely confirm the conclusion [2] that BII is a distinct B-form with differences in the values of three backbone torsions, namely ξ being in trans region, high value of ε and low value of β in comparison to BI conformations. We have showed, at least for some base sequences, that BII form corresponds to a local energy minimum on dDMP potential energy surface, and that the BII conformation is not a result of interactions with other molecules in the crystals. The conformation of dCpdG corresponding to BII-minimum of energy together with such conformation in duplex crystal [33] is shown in Fig. 4. The conclusion about the predisposed contribution of intra-chain interactions to a particular duplex conformation depending on specific base sequence will be further studied by computation of various energy minima on the potential energy surface of dDMPs. Such study is currently in progress.

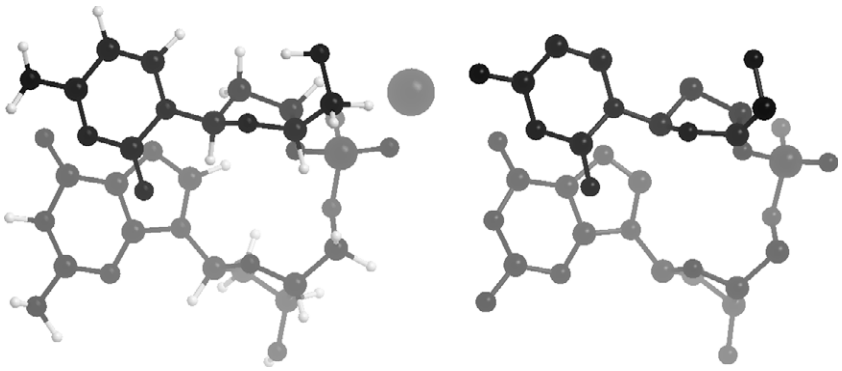

Fig. 4. The minimum energy conformation of the dCpdG–Na^+ complex corresponding to BII conformation (left), and the same step of d(CpTpCpTpCpGpApGpApG) duplex in the crystal [26] (right).

References

[1] H.M. Berman, W.K. Olson, D.L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S.-H. Hsieh, A.R. Srinivasan, B. Schneider, Biophys. J. 63 (1992) 751–759.
[2] D. Svozil, J. Kalina, M. Omelka, B. Schneider, Nucleic Acids Res. 36 (2008) 3690–3706.
[3] D.L. Beveridge, G. Barreiro, K.S. Byun, D.A. Case, T.E. Cheatham III, S.B. Dixit, E. Giudice, F. Lankas, R. Lavery, J.H. Maddocks, R. Osman, E. Seibert, H. Sklenar, G. Stoll, K.M. Thayer, P. Varnai, M.A. Young, Biophys. J. 87 (2004) 3799.
[4] A. Pérez, F.J. Luque, M. Orozco, J. Am. Chem. Soc. 129 (2007) 14739.
[5] D. Svozil, J.E. Šponer, I. Marchan, A. Pérez, T.E. Cheatham III, F. Forti, F.J. Luque, M. Orozco, J. Šponer, J. Phys. Chem. B 112 (2008) 8188.
[6] V.I. Poltev, V.M. Anisimov, V.I. Danilov, A. Deriabina, E. Gonzalez, A. Jurkiewicz, A. Leś, N. Polteva, J. Biomol. Struct. Dyn. 25 (2008) 563.
[7] C. Sosa, J. Andzelm, B.C. Elkin, E. Wimmer, K.D. Dobbs, D.A. Dixon, J. Phys. Chem. 96 (1992) 6630.
[8] N. Godbout, D.R. Salahub, J. Andzelm, E. Wimmer, Can. J. Chem. 70 (1992) 560.
[9] J.P. Perdew, Y. Wang, Phys. Rev. B 45 (1992) 13244.
[10] T. Tsuzuki, H.P. Luthi, J. Chem. Phys. 114 (2001) 3949.
[11] N. Kurita, H. Inoue, H. Sekino, Chem. Phys. Lett. 370 (2003) 161.
[12] T. van der Wijst, C.F. Guerra, M. Swart, F.M. Bickelhaupt, Chem. Phys. Lett. 426 (2006) 415.
[13] J. Šponer, P. Jurecka, P. Hobza, J. Am. Chem. Soc. 126 (2004) 10142.
[14] M. Swart, T. van der Wijst, C.F. Guerra, F.M. Bickelhaupt, J. Mol. Model. 13 (2007) 1245.
[15] P.C. Hariharan, J.A. Pople, Theor. Chim. Acta (Berlin) 28 (1973) 213.
[16] A.D. Becke, J. Chem. Phys. 98 (1993) 5648.
[17] M.W. Schmidt et al., J. Comp. Chem. 14 (1993) 1347.
[18] M.J. Frisch et al., Gaussian 03, Revision C.02, Gaussian Inc., Wallingford CT, 2004.
[19] G. te Velde, F.M. Bickelhaupt, S.J.A. van Gisbergen, C.F. Guerra, E.J. Baerends, J.G. Snijders, T. Ziegler, J. Comput. Chem. 22 (2001) 931.
[20] G. Sun, J.H. Voigt, I.V. Filippov, V.E. Marquez, M.C. Nicklaus, J. Chem. Inf. Model. 44 (2004) 1752.
[21] L. Clowney, S.C. Jain, A.R. Srinivasan, J. Westbrook, W.K. Olson, H.M. Berman, J. Am. Chem. Soc. 118 (1996) 509.
[22] A. Gelbin, B. Schneider, L. Clowney, S.H. Hsieh, W.K. Olson, H.M. Berman, J. Am. Chem. Soc. 118 (1996) 519.
[23] C. Altona, M. Sundaralingam, J. Am. Chem. Soc. 94 (1972) 8205.
[24] S. Fuberg, Acta Cryst. 3 (1950) 325.
[25] S. Wang, H.F. Schaefer III, J. Chem. Phys. 124 (2006) 044303.
[26] W. Koch, M.C. Holthausen, A Chemist's Guide to Density Functional Theory, second ed., Wiley, 2001.
[27] J. Šponer, J. Leszczynski, P. Hobza, Biopolymers 61 (2002) 3.
[28] R.E. Dickerson, in: R.H. Sarma, M.H. Sarma (Eds.), Structure, Motion, Interaction and Expression of Biological Macromolecules, vol. 1, Adenine Press, Albany, NY, 1998, pp. 17–36.
[29] K.K. Woods, L. McFail-Isom, C.C. Sines, S.B. Howerton, R.K. Stephens, L.D. Williams, J. Am. Chem. Soc. 122 (2000) 1546.
[30] H. Yuan, J. Quintana, R.E. Dickerson, Biochemistry 31 (1992) 8009.
[31] N. Valls, G. Wright, R.A. Steiner, G.N. Murshudov, J.A. Subirana, Acta Crystallogr. Sect. D 60 (2004) 680.
[32] G.G. Privé, K. Yanagi, R.E. Dickerson, J. Mol. Biol. 217 (1991) 177.
[33] D.S. Goodsell, K. Grzeskowiak, R.E. Dickerson, Biochemistry 34 (1995) 1022.

Bioorganic Chemistry 37 (2009) 65–69

Contents lists available at ScienceDirect

Bioorganic Chemistry

journal homepage: www.elsevier.com/locate/bioorg

Preliminary Communications

Synthesis and NMR properties of novel 5,6-dihydroborauracil derivatives

Tomasz Ruman [a,*], Karolina Długopolska [a], Anna Kuśnierz [a], Agata Jurkiewicz [b], Andrzej Leś [b,c], Wojciech Rode [a,d]

[a] Rzeszów University of Technology, Faculty of Chemistry, Department of Biochemistry and Biotechnology, 6 Powstańców Warszawy Avenue, Rzeszów 35-959, Poland
[b] University of Warsaw, Faculty of Chemistry, Quantum Chemistry Laboratory, 1 Pasteur Street, Warsaw 02-093, Poland
[c] Pharmaceutical Research Institute, Rydygier Street 8, Warsaw 01-793, Poland
[d] Nencki Institute of Experimental Biology, 3 Pasteur Street, Warsaw 02-093, Poland

ARTICLE INFO

Article history:
Received 4 February 2009
Available online 24 March 2009

Keywords:
Borauracil
Dihydroborauracil
Hydroboration
N-vinylurea

ABSTRACT

Novel boron compounds – 5,6-saturated borauracil derivatives (4-bromo-5,6-dihydroborauracil, 4-hydroxy-5,6-dihydroborauracil and 4-methoxy-5,6-dihydroborauracil) are presented along with other boron compounds obtained from *N*-vinylurea: *N*-substituted β-boronic amino acid – 2-{[(dihydroxyboranoamino)(dihydroxyboranooxy)methyl]-amino}ethylboronic acid and substituted methoxy-borane O-[(1-amino-1-*N*-vinylamino)methyl]dihydroxyboronate.

1. Introduction

Hydroboration, an important reaction in synthetic organic chemistry, is accomplished with various hydroborating agents, including borane-tetrahydrofuran, borane-pyridine, borane-dimethyl sulfide, 9-BBN, and thexylborane complexes. Those reagents are commercially available and offer many options for selective hydroboration [1–5].

Little is known about the hydroboration of olefinic ureas compared to other olefinic systems; the published data have been reviewed by Soloway et al. [6,7]. Only a few boronic acids containing urea-derived moieties are known and the intermediates formed during the respective reactions were not described.

In this paper we present novel boron compounds **1**–**6**. Compounds **1** and **2**, i.e. a substituted methoxyborane (*O*-[(1-amino-1-*N*-vinylamino)methyl]dihydroxyboronate), and *N*-substituted β-boronic amino acid, respectively (Fig. 1), could be used as building blocks for pharmaceuticals or as cross-linking agents in polymer chemistry.

The most important aspect of this work are novel boron analogues of 5,6-dihydrouracil (5,6-dihydroborauracils, **4**–**6** see Fig. 2). There are only few boron analogues of nucleic acid bases, with Zhuo group's benzoborauracils as the most important work of the last 20 years [8]. It should be noted that borauracils, borathymines or boracytosines and their derivatives should be key compounds in the search for e.g. new anticancer compounds. Boron analogues of biologically active compounds have very interesting properties. The mechanism of action that is unique for those compounds is believed to be the nucleophilic attack of an enzyme on to the electron deficient sp^2 boron, resulting in the tetrahedral sp^3 hybridized boron "ate" complex. The boron analogue–enzyme "ate" complex is highly stable resulting in effective inhibition of the enzyme [9–11]. Additionally, boron-containing compounds can be considered potentially active in boron–neutron capture therapy (BNCT) [12].

2. Results and discussion

2.1. Hydroboration of N-vinylurea

Of particular interest is the fact that *O*-[(1-amino-1-*N*-vinylamino)methyl]dihydroxyboronate (**1**) was synthesized in the present study by an almost identical method to that used by Soloway's group [12]. Instead of the expected hydroboration product (borane to C=C double bond addition product), compound **1** is the product of addition at the carbonyl group of the vinylurea (Fig. 1). While our synthetic approach allowed the substrates, *N*-vinylurea and borane complex ($BH_3 \cdot THF$), to be used in strictly equimolar quantities, the same reaction would have been difficult to perform using gaseous diborane as reported [6,7,13].

A proposed intermediate compound **1′** resulting from addition of borane to the carbonyl group of *N*-vinylurea undergoes hydrolysis to give **1**. The vinyl moiety of *N*-vinylurea was not affected during the reaction according to NMR spectroscopy. Carbonyl group reduction by diborane gas and borane complexes have been

* Corresponding author.
E-mail address: tomruman@prz.edu.pl (T. Ruman).

0045-2068/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.bioorg.2009.03.003

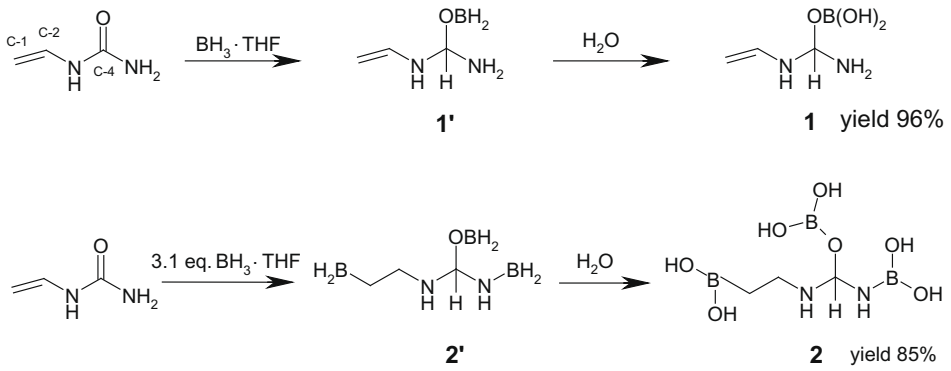

Fig. 1. Synthesis of compounds **1** and **2**.

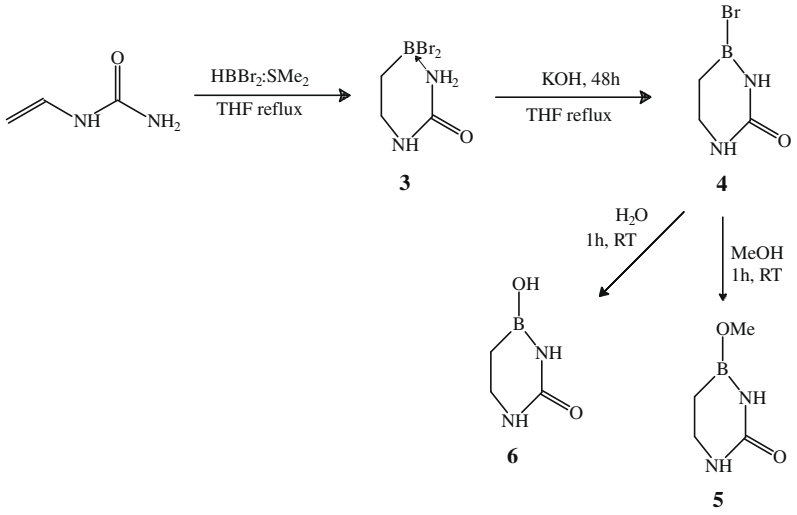

Fig. 2. Synthetic route for preparation of compounds **3**–**6**.

studied widely by both experimental and computational approaches [14–16]. The reduction of urea moiety should not be surprising as boranes also are known to perform reduction of wide range of amides. The typical reduction of amide is performed in mild conditions (66 °C, atmospheric pressure). Reduction with boranes in these conditions can lead even to amines [17]. The proposed reaction mechanism suggests that the methine hydrogen atom originates from the borane complex, with the remaining BH_2 moiety being attached to the oxygen atom. The final dihydroxyboron structural fragment then results from hydrolysis.

The presented experimental data shows that hydroboration of the unsaturated vinyl moiety also resulted in the formation of proposed intermediate **2′**, which could be easily hydrolyzed to dihydroxyboron derivative **2**. When the borane to *N*-vinylurea molar ratio was $\geqslant$1.8, addition of borane to the carbonyl moiety took place, followed by hydroboration of the olefinic fragment and substitution of hydrogen on the peripheral $-NH_2$ group (Fig. 1), giving **2′**, which hydrolyzed to compound **2**. Using borane to vinylurea molar ratios ranging from 1.8 to 2.8 resulted in formation of compounds **1** and **2** in different proportions. Compound **2** was obtained as the major product when the borane to vinylurea molar ratio was $\geqslant$3.1. No products containing two dihydroxyboron moieties were detected. Surprisingly, our synthetic route to **1** is very similar to that previously described, leading to β-ureidoethylboronic acid by direct C=C bond hydroboration, thus having an unreacted carbonyl moiety of boronic acid [6,7]. It should be noted that newly formed *O*-linked boric acid esters (**1** and **2**) could react with water and also with simple alcohols and due to this reactivity the solutions of those compounds for NMR spectra were prepared in aprotic solvents.

2.2. Computational analysis of the regio- and chemoselectivity of N-vinylurea and its derivatives

The alkene and alkyne hydroboration reaction mechanism is well known.[17] Steric effects are responsible for the regioselectivity, especially when using bulky borane reagents, however, for the majority of olefins, the most significant factors are electronic effects. Attack of the electrophilic reagent, e.g. BH_3, causes significant redistribution of the electron density in the target molecule, and as a consequence affects its reactivity. In the present work we considered a simplified mechanism for the hydroboration of *N*-vinylurea by borane-tetrahydrofuran reagent.

The atomic charges calculations show that the sum of the natural charges of the olefinic moiety (C-1, C-2, and adjacent hydrogens) in *N*-vinylurea is more positive (+0.153 a.u.) than in intermediate **1′** (+0.098 a.u.), indicating that *N*-vinylurea should react with the borane complex more slowly than **1′**. This can be rationalized by the lack of a carbonyl-related strong resonance effect in **1** and **1′** which is known for amides and also occurs in urea derivatives. The calculated natural charge on the C-4 carbonyl atom, is +0.771 a.u., and this is the highest positive partial charge found

for *N*-vinylurea. Therefore, it is reasonable to expect that the partially negative hydrogen atom of the borane complex (for comparison, the natural charge of the H atom in BH_3 is −0.107 a.u.) should attack the carbonyl carbon [18–20]. This suggestion agrees with the experimental results. In summary, compounds **2** and **2′** can be obtained from *N*-vinylurea *via* hydroboration of compound **1′** which is a borane-carbonyl addition product (*vide supra*).

2.3. Synthesis of 5,6-dihydroborauracils

The knowledge of small borane (BH_3 and its complexes) reactivity towards carbonyl and amide groups suggested the use of the bulky dibromoborane complex for the reaction with vinylurea. Experimental data proved that dibromoborane–dimethylsulfide complex does not react with carbonyl group as it was observed for BH_3:SMe_2 complex. The possible explanation for this phenomenon is lower reactivity of $BHBr_2$ reagent caused by both much higher steric repulsion and a strong electron withdrawing effect from two bromine atoms. Moreover, tetrahydrofuran could easily form $BHBr_2$:THF complex with even higher steric hindrance. The $BHBr_2$ complex hydroboration step (Fig. 2) was conducted at much higher temperature (66 °C) than in the case of BH_3 complex (0 °C) and unfortunately it resulted in slow loss of dibromoborane. This negative effect was neutralized by multiple additions of dibromoborane complex every hour of reaction time.

The obtained linear product **3** (Fig. 2) is similar to the one obtained by Soloway group having BBr_2 instead of BH_2 group [12]. The work mentioned suggested possibility of formation of six-membered complex where carbonyl oxygen coordinates to boron atom. Such a possibility also exists for compound **3** with coordination of carbonyl oxygen or, more likely, peripheral nitrogen atom (Fig. 2). The same group obtained also a linear boronic acid $(HO)_2BCH_2CH_2NHCONH_2$ but failed to convert it into cyclic product with covalent B–N bond [13].

The reaction of **3** with potassium hydroxide gave cyclic six-membered product 4-bromo-5,6-dihydroborauracil (**4**). To the best of our knowledge, **4** is the first boron-containing 5,6-saturated analogue of uracil. The formation of covalent B–N bond was confirmed by our analytical data, but unfortunately, we were unable to prepare X-ray diffraction quality monocrystal. The ^{11}B NMR chemical shift of compound **4** (54.5 ppm) confirmed that boron atom is covalently connected to carbon, nitrogen and bromine atoms [21–23]. The exceptionally interesting aspect is the possibility of coordination to boron atom of chiral electron pair donors (e.g. lactate, amino acids, etc.) that should result in formation of diastereoisomeric products. Another dihydrouracils – 4-*O*-methoxy-5,6-dihydroborauracil (**5**) and 4-hydroxy-5,6-dihydroborauracil (**6**) were prepared by methanolysis (**5**) or hydrolysis (**6**) of **4**. The ^{11}B chemical shifts of **5** (23.7 ppm) and **6** (24.9 ppm) are in accord with literature data [8,21–23]. There was also observed the formation of tetrahedral bis-methanol adduct of **5** (**5**:CD_3OD) with ^{11}B chemical shift of 1.6 ppm, similar to those described by Zhuo et al. [8]. The problem of determining the atom bound to boron was previously discussed by Zhuo et al. [8]. The authors proved that hydroxyboron compound (B–OH cyclic product) of benzoborauracils contains B–N bond in solid state or in non-coordinating solvent, while the titration with methanol gave bismethanol adduct that may contain B–O (carbonyl oxygen coordinating to boron) form in equilibrium with B–N form in solution. The removal of methanol gave starting B–N form as judged from ^{11}B NMR spectra [8]. Our observations are very similar to that of Zhuo et al. Solution of compound **6** in non-coordinating solvent contains mainly B–N form with ^{11}B chemical shifts value of 24.9 ppm (for **6**), a value that differs significantly from literature data for alkylboronic acids and their esters ((R–$B(OH)_2$ and R–$B(OR)_2$, respectively, R=alkyl) where chemical shifts that are always in 30–34 ppm range [24–28]. It should be noted that methanol (or other alcohol of small steric hindrance) solution could contain some amount of B–O form, the quantitative analysis of B–O/B–N equilibrium is difficult due to process rate that is fast in NMR-scale in analyzed conditions (0–20 °C).

3. Conclusions

The synthesis, reactivity and spectroscopic properties of novel boron compounds are presented. The intermediate compound **1′** is the product of addition of the borane complex at the carbonyl group of *N*-vinylurea giving *O*-[(1-amino-1-*N*-vinylamino)methyl]dihydroxyboronate (**1**) after hydrolysis. No reaction in the vinyl moiety of *N*-vinylurea was apparent under conditions used. Borane to *N*-vinylurea molar ratios higher than 1.8 resulted in addition of borane to the carbonyl moiety, followed by hydroboration of the olefinic fragment and substitution of a hydrogen from the peripheral–NH_2 group. Borane to vinylurea molar ratios ranging from 1.8 to 2.8 resulted in the formation of compounds **1** and **2** in different proportions. The calculated natural charge on the C-4 carbonyl atom of +0.771 a.u. is the highest positive partial charge found for *N*-vinylurea. Therefore, the partially negative hydrogen atom from the borane complex may be expected to attack the carbonyl carbon, in accord with the experimental results.

Reaction of dibromoborane complex with *N*-vinylurea conducted in higher temperature (66 °C) resulted in synthesis of C=C double bond hydroboration product **3**, formally a boronic acid bromide. Its reaction with potassium hydroxide in aprotic solvent resulted in heterocyclic product **4**, a first 5,6-dihydroborauracil derivative. Its hydrolysis gave B–OH moiety at 4-position of saturated ring (**6**), and reaction with methanol resulted in formation of another cyclic compound containing B–OMe structural fragment (**5**).

4. Experimental

4.1. Materials and methods

^1H NMR spectra were obtained with Bruker Avance spectrometer operating in the quadrature mode at 500 MHz. All ^{11}B spectra were performed using 5 mm pure quartz NMR tube. The residual peaks of deuterated solvents were used as internal standards. Elemental analysis was performed using Carlo Erba Elemental Analyser EA 1108. GC–MS analysis was carried out on Agilent Technologies 6890 N apparatus with 5973-Network mass detector. FTIR spectra was recorded on Perkin Elmer Paragon 1000 apparatus. The *N*-vinylurea was prepared using known method [29]. All other reagents and deuterated solvents of the highest commercially available grade were purchased from Aldrich and used without further purification (with exception for acetone-*d*6 which was dried with anhydrous sodium sulfate). Rubber septa joints were purchased from Aldrich. All procedures, including preparation of samples for the NMR measurements, were carried out under nitrogen.

The theoretical calculations have been performed with the density functional theory and the B3LYP functional with the aug-cc-pVTZ basis set implemented in the Gaussian G03 (rev. C.02) suite of programs [30]. The optimal geometries were obtained with default methods achieving the stationary point controlled by four convergence criteria. The electronic density distribution has been analyzed with the use of the NBO 5.G module implemented also in the Gaussian G03.

4.2. Preparation of O-[(1-amino-1-N-vinylamino)methyl]-dihydroxyboronate (1)

1 M Borane–tetrahydrofuran complex in THF (1 g of borane–THF complex, 11.6 mmol) was placed in a dry, 100 ml round-

bottom flask purged with dry nitrogen. The flask was then immersed in an ice bath. *N*-vinylurea [29] (1 g in 5 ml of dry THF, 11.6 mmol) was added dropwise over a period of 30 min and the reaction was allowed to run for another 30 min, with the reaction temperature rising to 5 °C. Distilled and deoxygenated water (630 μL, 35 mmol) was added dropwise. Following completion of hydrogen evolution, all volatile products were removed under reduced pressure. Finally, the resulting solid was dried under high vacuum. Compound **1** was obtained in 96% yield (*ca.* 1.46 g). MS (electrospray ionization, *m*/*z*): 131 + 132 (100%). Elemental Anal. Calcd for $C_3H_9BN_2O_3$: C, 27.31; H, 6.83; N, 21.24. Found: C, 27.21; H, 6.86; N, 21.22. FTIR (KBr, cm^{-1}) 3320 (νN–H, s, br); 3223 (νB–O–H, s, br); 2104 (νC═CH, w); 1663 (νN–H, m); 1617 (νC═C, m); 1560 (νC–O, m); 1345 (νB–O, m); 1026 (νC–N, m); 750–760 (νN–H, w). ^{1}H NMR (acetone-*d*6, ppm): 7.67–7.78 (m, H_{vinyl}, 3H); 5.77 (s, OH, 1H), 3.64 (m, CH, 1H), 1.72 (m, NH, 1H); 1.46 (m, NH_2, 2H). ^{13}C NMR (acetone-*d*6, ppm): 83.2 ($CH_{satur.}$); 115.8, 165.0 (C_{vinyl}). ^{11}B NMR (acetone-*d*6, ppm): 19.21.

4.3. Preparation of 2-{[(dihydroxyboranoamino) (dihydroxyboranooxy)methyl]amino}ethylboronic acid (*2*) (Fig. 1)

1 M Borane–tetrahydrofuran complex in THF (3.1 g of borane-THF complex, 36.1 mmol) was placed in a dry, 100 ml round-bottom flask purged with dry nitrogen. The flask was then immersed in an ice bath. *N*-vinylurea [29] (1 g, 11.6 mmol, in 5 ml of dry THF) was added dropwise over a period of 30 min, and the reaction was run for another 90 min, with the reaction temperature rising to 10 °C. Distilled and deoxygenated water (2 ml, 0.108 mol) was added dropwise. Following completion of hydrogen evolution, all the volatile products were removed under reduced pressure and the resulting solid was dried under high vacuum. Compound **2** was obtained in 85% yield (*ca.* 2.2 g). MS (electrospray ionization, *m*/*z*): 219–222 (100%). Elemental Anal. Calcd for $C_3H_{13}BN_2O_7$: C, 16.26; H, 5.87; N, 12.65. Found: C, 16.22; H, 5.90; N, 12.57. ^{1}H NMR (acetone-*d*6, ppm): 7.0, 7.27, 5.30 (br., OH, 2H + 2H + 2H) 3.30 (m, CH, 1H); 3.13 (m, C*H*$_2$–N, 2H); 2.18 (m, N*H*–CH_2, 1H); 2.14 (d, N*H*–B, 1H) 1.06 (tr., B–C*H*$_2$, 2H). ^{11}B NMR (acetone-*d*6, ppm): 20.24; 19.70, −13.02. Compounds **1** and **2** were also obtained by analogous reactions of *N*-vinylurea with 1 M toluene solution of borane–dimethylsulfide complex, in yields of 92% and 79%, respectively.

4.4. Preparation of *3* and *4*

Dibromoborane–dimethylsulfide complex (1.5 g of dibromoborane–Me_2S complex, 6.4 mmol) was placed in a 100 ml round-bottom flask containing 20 ml of anhydrous THF. The flask was then immersed in a silicon oil bath. *N*-vinylurea [29] (550 mg in 5 ml of dry THF) was added dropwise to the reaction flask at THF reflux conditions over a period of 10 min and the reaction was allowed to run for another 30 min. Other portions of dibromoborane–dimethylsulfide complex (1.5 g) were added every 1 h to the refluxing mixture (total amount of $BHBr_2$:SMe_2 complex used in the reaction was 15 g). All volatile products were removed under reduced pressure. Finally, the resulting solid was dried under high vacuum. Compound **3** was obtained in 85% yield as judged from NMR spectra. Reaction product was dissolved in anhydrous THF (20 ml) containing 3.6 g of dry KOH and refluxed for 48 h. Resulting product was vacuum dried. Compound **4** was washed out with two portions of chloroform (2 × 10 ml). Yield: 880 mg (78% total yield). Analytical data for **4**: MS (electrospray ionization, *m*/*z*): 176–177 (100%). Elemental Anal. Calcd for $C_3H_6BBrN_2O$: C, 20.38; H, 3.42: N, 15.84. Found: C, 20.30; H, 3.46; N, 15.81. ^{1}H NMR (acetone-*d*6, ppm): 5.80–5.56 (br, NH, 2H); 3.35 (m, N–C*H*$_2$, 2H), 1.18 (m, B–C*H*$_2$, 2H). ^{13}C NMR (acetone-*d*6, ppm): 155.2 (C═O); 35.1 (N–C*H*$_2$), 15.6 (B–C*H*$_2$). ^{11}B NMR (acetone-*d*6, ppm): 54.5.

4.5. Preparation of *5*

To the solution of **4** in anhydrous tetrahydrofurane (200 mg in 10 ml of THF) anhydrous methanol was added (50 ml). Reaction mixture was stirred in room temperature for 1 h and vacuum dried (60 °C, 48 h at 0.01 Pa). Compound **5** was obtained in almost quantitative yield (approx. 0.14 g). Analytical data for **5**: MS (electrospray ionization, *m*/*z*): 127–128 (100%). Elemental Anal. Calcd for $C_4H_9BN_2O_2$: C, 37.55; H, 7.09; N, 21.90. Found: C, 37.48; H, 7.11; N, 21.84. ^{1}H NMR (acetone-*d*6, ppm): 5.61–5.72 (br, NH, 2H); 3.61 (s, OMe, 3H), 3.38 (m, N–CH_2, 2H), 1.28 (m, B–CH_2, 2H). ^{13}C NMR (acetone-*d*6, ppm): 157.2 (C═O); 53.1 (OMe); 35.2 (N–CH_2), 17.0 (B–CH_2). ^{11}B NMR (acetone-*d*6, ppm): 23.7. ^{11}B NMR (methanol-d4, ppm): 1.6.

4.6. Preparation of *6*

To the solution of **4** in anhydrous tetrahydrofurane (300 mg in 10 ml of THF) deoxygenated water was added (10 ml). Reaction mixture was stirred in room temperature for 1 h and vacuum dried. Compound **6** was obtained in almost quantitative yield (approx. 0.19 g). Analytical data for **6**: MS (electrospray ionization, *m*/*z*): 113–114 (100%). Elemental Anal. Calcd for $C_3H_7BN_2O_2$: C, 31.6; H, 6.19; N, 24.59. Found C, 31.54; H, 6.23; N, 24.52. ^{1}H NMR (acetone-*d*6, ppm): 8.43 (br, B–OH, 1H); 5.68–5.81 (br, NH, 2H); 3.42 (m, N–CH_2, 2H), 1.31 (m, B–CH_2, 2H). ^{13}C NMR (acetone-*d*6, ppm): 156.2 (C═O); 35.0 (N–C*H*$_2$), 17.5 (B–C*H*$_2$). ^{11}B NMR (acetone-*d*6, ppm): 24.9. ^{1}H NMR (D_2O, ppm): 3.48 (m, N–CH_2, 2H), 1.34 (m, B–CH_2, 2H).

Acknowledgement

This work was supported by the Ministry of Education and Science, Poland, Grant No. N204 088 31/2052 2006–2009. The theoretical calculations were performed in the ICM UW Warsaw University computer center within the G18-6 computer grant.

References

[1] J.M. Clay, E. Vedejs, J. Am. Chem. Soc. 127 (2005) 5766–5767.
[2] K. Shirakawa, A. Arase, M. Hoshi, Synthesis (2004) 1814–1820.
[3] J.V.B. Kanth, H.C. Brown, J. Org. Chem. 66 (2001) 5359–5365.
[4] G.W. Kabalka, T.M. Shoup, N.M. Goudgaon, J. Org. Chem. (1989) 5930–5933.
[5] E. Negishi, H.C. Brown, Synthesis (1974) 77–89.
[6] D.N. Butler, A.H. Soloway, J. Am. Chem. Soc. 88 (3) (1966) 484–487.
[7] V.M. Dembitsky, G.A. Tolstikov, M. Srebnik, Eurasian Chem. Tech. J. 4 (2002) 153–167.
[8] J.-C. Zhuo, A.H. Soloway, J.C. Beeson, W. Ji, B.A. Barnum, F.-G. Rong, W. Tjarks, G.T. Jordan IV, J. Liu, S.G. Shore, J. Org. Chem. 64 (1999) 9566–9574.
[9] V.J. Reddy, S. Chandra, M.V. Ram Reddy, Org. Biomol. Chem. 5 (2007) 889–891.
[10] W. Yang, X. Gao, B. Wang, Med. Res. Rev. 23 (2003) 346–368.
[11] R.R. Wolfenden, Annu. Rev. Biophys. Bioeng. 5 (1976) 271.
[12] A.H. Soloway, W. Tjarks, B.A. Barnum, F.-G. Rong, R.F. Barth, I.M. Codogni, J.G. Wilson, Chem. Rev. 98 (1998) 1515–1562.
[13] D.N. Butler, A.H. Soloway, J. Med. Chem. 9 (3) (1966) 362–365.
[14] M. DiMare, J. Org. Chem. 61 (1996) 8378–8385.
[15] D.-M. Du, T. Fang, J. Xu, S.-W. Zhang, Org. Lett. 8 (2006) 1327–1330.
[16] J.W. Bae, S.H. Lee, Y.J. Jung, C.-O. Maing, C.M. Yoon, Tetrahedron Lett. 42 (2001) 2137–2139.
[17] R.W. Roeske, F.L. Weitl, K.U. Prasad, R.M. Thompson, J. Org. Chem. 41 (7) (1976) 1260.
[18] H.C. Brown, G. Zweifel, J. Am. Chem. Soc. 83 (1961) 2544–2551.
[19] R.G. Parr, W. Yang, J. Am. Chem. Soc. 106 (1984) 4049–4050.
[20] W. Yang, W.J. Mortier, J. Am. Chem. Soc. 108 (1986) 5708–5711.
[21] G.R. Eaton, J. Chem. Educ. 46 (1969) 547.
[22] W.L. Smith, J. Chem. Educ. 54 (1977) 469.
[23] S. Hermanek, Chem. Rev. 92 (1992) 325.
[24] J.E. DeMoor, G.P. Van der Kelen, J. Organomet. Chem. 6 (1966) 235.
[25] W.D. Phillips, H.C. Miller, E.L. Mutterties, J. Am. Chem. Soc. 81 (1959) 4496.
[26] H. Nöth, H. Vahrenkamp, Chem. Ber. 99 (1966) 1049.

[27] H.C. Brown, J.B. Campbell Jr, J. Org. Chem. 45 (1980) 389.
[28] K. Maruyama, J. Org. Chem. 42 (1977) 3252.
[29] R. Hart, Bull. Soc. Chim. Belg. 66 (1957) 229–243.
[30] Gaussian 03, Revision C.02, M.J. Frisch, Trucks, H.B. Schlegel, G. E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery, Jr., T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople, Gaussian, Inc., Wallingford CT, 2004. (http://www.gaussian.com and references therein).

Available online at www.sciencedirect.com
ScienceDirect
Inorganic Chemistry Communications 10 (2007) 1074–1078
INORGANIC CHEMISTRY COMMUNICATIONS
www.elsevier.com/locate/inoche

The synthesis, reactivity and ^{1}H NMR investigation of the hydroxyborohydride anion

Tomasz Ruman [a,*], Anna Kuśnierz [a], Agata Jurkiewicz [b], Andrzej Leś [b], Wojciech Rode [c]

[a] *Rzeszów University of Technology, Faculty of Chemistry, Department of Biochemistry and Biotechnology, 6 Powstańców Warszawy Avenue, 35-959 Rzeszów, Poland*
[b] *University of Warsaw, Faculty of Chemistry, Quantum Chemistry Laboratory, 1 Pasteur Street, 02-093 Warsaw, Poland*
[c] *Nencki Institute of Experimental Biology, 3 Pasteur Street, 02-093 Warsaw, Poland*

Received 15 May 2007; accepted 29 May 2007
Available online 8 June 2007

Abstract

Hydroxyborohydride (sodium and potassium salts) were prepared by a one-pot method and the ^{1}H NMR spectrum of sodium hydroxyborohydride was recorded for the first time. Based on an NMR experiment, hydroxyborohydride was found to be more reactive toward acetone carbonyl than borohydride. The calculated ^{11}B- and ^{10}B–H coupling constants are in very good agreement with experimental data, and support the hypothesis about the detection of the $(BH_3OH)^-$ anion as an individual species in solution.

Keywords: Boron; ^{1}H NMR; Hydroxyborohydride; Reduction; Carbonyl compounds

Introduction

Both organic and inorganic boron compounds have been widely used in synthetic chemistry, their development being fast [1,2], particularly in the last fifteen years, e.g.: Suzuki reagents [3,4], Soderquist reagents [5–7], boryllithium [1] and many others [8–16]. The most popular inorganic boron reagents used in chemical synthesis are borohydride salts that are especially suitable for reduction of carbonyl species.

The hydroxyborohydride ion $(BH_3OH)^-$ is a very useful reducing agent, used usually for the reduction of carbonyl compounds with limited solubility in organic solvents of low polarity [17]. Reduction of carbonyl compounds by BH_4^- (borohydride), compared to $(BH_3OH)^-$ (hydroxyborohydride), shows usually higher reaction rates, and occurs at significantly lower temperatures [18].

Hydroxyborohydride ion is formed during hydrolysis of the borohydride ion BH_4^- under neutral or acidic conditions. Very few approaches have been described, including those utilizing flow reactors, to synthesize hydroxyborohydride salts. Surprisingly, the reaction was, in most cases, carried out in aqueous solutions [19–21].

A new, fast, one-pot method of sodium/potassium hydroxyborohydride preparation is presented, along with the ^{1}H NMR spectrum of hydroxyborohydride, recorded for the first time. Additionally, NMR studies demonstrated higher reactivity of hydroxyborohydride than borohydride toward acetone carbonyl.

Experimental

Methods and materials

^{1}H NMR spectra were obtained with Bruker Avance spectrometer operating in the quadrature mode at 500 MHz. The residual peaks of deuterated solvents were used as internal standards. Elemental analysis was performed using Carlo Erba Elemental Analyser EA 1108. GC–MS analysis was carried out on Agilent Technologies 6890N apparatus with 5973-Network mass detector. All

* Corresponding author. Tel.: +48178651759.
E-mail address: tomruman@prz.edu.pl (T. Ruman).

1387-7003/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.inoche.2007.05.029

reagents and deuterated solvents of the highest commercially available grade were purchased from Aldrich, and used without further purification. Rubber septa joints were purchased from Aldrich. All procedures, including preparation of samples for the NMR measurements, were carried out under nitrogen.

Theoretical calculations were performed with the B3LYP/aug-cc-pVTZ method and, in a few cases, also with the PW91PW91/aug-cc-pVTZ method. The optimal geometries were obtained and confirmed with positive harmonic frequencies. All calculations were performed with the Gaussian G03 (Rev. C.02) suite of programs [22–30]. The pictures of calculated structures were made with the Molden program and are presented in Supplementary materials [31].

Synthesis of sodium hydroxyborohydride

Sodium borohydride (0.462 g, 12.2 mmol) was placed in a dry, 100 ml round-bottom flask, containing 20 ml anhydrous tetrahydrofurane and a magnetic stirring bar, fitted with rubber septa, and immersed in a water bath at room temperature. Deoxygenated water (0.22 ml, 12.2 mmol) was added dropwise. The progress of the reaction was associated by hydrogen sodium borohydride particles being dissolved, and the solution becoming slightly turbid. Following completion of hydrogen evolution, all volatile products were removed under reduced pressure. Finally, the solid was dried under high vacuum. Sodium hydroxyborohydride was obtained in 96% yield (ca. 0.63 g). GC–MS studies gave uncertain results due to high reactivity of the hydroxyborohydride ion, $m/z = 54$ (100%). Elemental analysis: H 7.51% (calculated 7.43%). ^{1}H NMR ($(CD_3)_2CO$, ppm): 2.96 (OH, broad, 1H); −0.46, −0.29, −0.13, 0.03 (BH, quartet, 3H total); −0.37, −0.32, −0.26, −0.21, −0.16, −0.10, −0.05 (B*H*, septet, 3H total). More detailed analysis of the spectrum is presented in ''Results and discussion''.

Synthesis of potassium hydroxyborohydride

The procedure was the same as that used for sodium hydroxyborohydride, except that potassium borohydride (0.501 g, 9.3 mmol) was a substrate and the resulting yield was 95%. Elemental analysis: H 5.65% (calculated 5.72%). The ^{1}H NMR spectrum of potassium hydroxyborohydride was identical to the one recorded for sodium hydroxyborohydride.

Reaction of sodium hydroxyborohydride or sodium borohydride with acetone-d6

Either sodium hydroxyborohydride (10 mg) or sodium borohydride (10 mg) was placed under nitrogen in an NMR tube sealed with a rubber Septa® (at 273 K). Acetone-*d*6 (0.5 ml; cooled to 273 K) was injected into the tube, the mixture vigorously shaken and ^{1}H NMR spectrum immediately recorded. Following 6 h incubation at 300 K, another spectrum was recorded.

The spectrum after 6 h reaction with hydroxyborohydride: ^{1}H NMR ($(CD_3)_2CO$): 3.92 (C*H*, s, 1H), 3.00 (O*H*, broad, 1H); −0.42, −0.26, −0.10, 0.07, (B*H*, quartet, 2H total); −0.34, −0.28, −0.23, −0.17, −0.12, −0.07, −0.01 (B*H*, septet, 2H total).

The spectrum after 6 h reaction with borohydride: No reaction between acetone-*d*6 and sodium borohydride was observed under such conditions: ^{1}H NMR ($(CD_3)_2CO$): −0.48, −0.32, −0.16, 0.00, (BH_4^-, quartet); −0.40, −0.35, −0.29, −0.24, −0.19, −0.13, −0.08 (BH_4^-, septet). $^1J_{H-11B} = 80.0$ Hz, $^1J_{H-10B} = 27.0$ Hz.

Results and discussion

Synthesis of hydroxyborohydride salts

There are very few known methods for synthesis of hydroxyborohydrides. All of them are based on the hydrolysis of borohydrides in aqueous solutions. Therefore, the resulting products include compounds from every step of borohydride hydrolysis, beside the final boric acid [19]. In contrast, the present synthetic route, carried out in an aprotic solvent, resulted in sodium and potassium hydroxyborohydrides in much higher yields.

The analysis of ^{1}H NMR spectrum of hydroxyborohydrides

The ^{1}H NMR spectra of hydroxyborohydride salts have not been previously published, probably due to synthetic problems and high reactivity towards atmospheric oxygen and moisture. The ^{1}H NMR spectrum of sodium hydroxyborohydride in acetone-*d*6 (Fig. 1) is similar to that recorded for sodium borohydride, showing similar B–H coupling region (approx. −0.5 to 0.0 ppm).

The most unusual and interesting part of this spectrum is the −0.5 to 0.1 ppm region containing B*H* resonances.

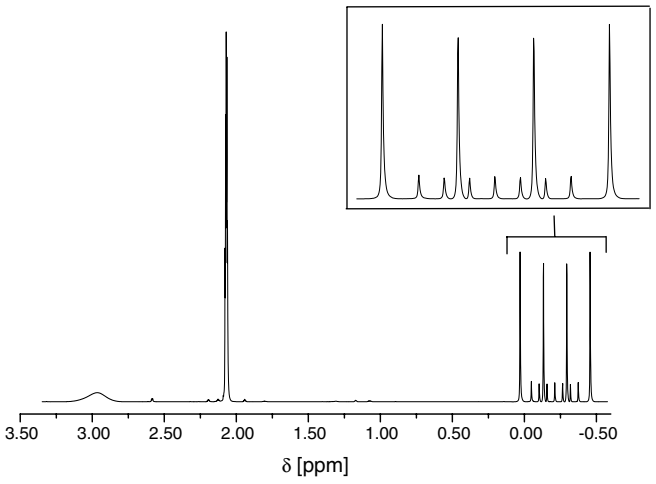

Fig. 1. ^{1}H NMR spectrum of sodium hydroxyborohydride in acetone-*d*6. The insert shows the −0.5 to 0.1 ppm region.

Natural boron contains two "NMR-active" isotopes: ^{10}B (with nuclear spin I and natural abundance of 3 and 19.9%, respectively) and ^{11}B (3/2 and 80.1%, respectively). The four peaks of larger intensity are due to heteronuclear coupling between ^{11}B and boron-attached hydrogen atoms. On the other hand, seven peaks with smaller intensity are the effect of coupling between ^{10}B and hydrogen atoms. The ^1H NMR spectrum of hydroxyborohydrides significantly differs from those recorded for borohydrides. The experiments conducted in identical conditions show noticeable chemical shift differences, allowing precise assignments. The measured coupling constants $^1J_{H-11B}$ were close to those found for borohydride anion (81.0 Hz) and the second isotope $^1J_{H-10B}$ (approx. 27.2 Hz). A similar value of $^1J_{H-11B}$ (82 Hz) was also found in a low quality ^{11}B NMR spectrum, presented in 1964 by Gardiner and Collat [19], and this result was also confirmed by our NMR experiments. The integration of ^{11}B- and ^{10}B-coupled hydrogen atom resonances is in agreement with the isotope composition of natural boron, and shows their relative intensities to be 4:1.

The reactivity of hydroxyborohydride ion

To our knowledge differences in reactivity between borohydride and hydroxyborohydride ions toward carbonyl compounds have been described only in a few cases [20,21]. The presented comparison of reducing power of both ions toward acetone-*d*6, under conditions of a large excess of the latter, was performed in an NMR-monitored experiment. The deuterated reagent was chosen due to the relative simplicity of the products spectra. With hardly any available data published on the subject, the mechanism of reduction of acetone-*d*6 by hydroxyborohydride ions may be expected to be similar to that observed for borohydride ions. Such a mechanism is also supported by the ^1H NMR spectrum recorded after 6 h reaction, containing an additional slightly broadened resonance at 3.92 ppm (half width 5.7 Hz), belonging to hydrogen atom bound to the secondary carbon of the isopropoxide-*d*6 moiety. The proton responsible for this signal seems to belong neither to hydroxyborohydride anion nor to acetone and the broadening of its resonance may be explained as due to coupling with deuterium atoms in the product isopropyloxy-*d*6 moiety. Thus the proton, to which the resonance was assigned, had to be transferred to the carbonyl carbon atom of acetone-*d*6 from the $(BH_3OH)^-$ molecule.

Calculation of optimal geometry and coupling constants

Several model systems have been chosen for comparison and for assessment of trends rather than attempting to reproduce absolute values of NMR parameters. Such a strategy stems from the fact that the calculation of spin–spin coupling constants is extremely sensitive to the choice of the molecular basis sets [32].

The results of $^1J(^{11}B–H)$ spin–spin coupling constants are presented in the Table 1. It can be seen that the values of $J(^{11}B–H)$ are close for BH_4^- and $(BH_3OH)^-$, but the values for $(BH_3OH)Na$ are different. In the latter case there is one value of $(^{11}B–H)$ at about 120 Hz, which does not correspond to the experimental spectrum. It seems reasonable to assume that the experimental value of J(B–H) should correspond to the averaged theoretical value because of rotation of the OH group. For example, taking a weighted average, we obtain a value of $(2 \cdot 79.7 + 84.6)/3 = 81.3$ Hz (PW91PW91), which corresponds quite well to the experimental value of 80.0 Hz. When the OH^- group is forced to be located on the three-fold axis of symmetry, i.e. when the B–O–H angle is fixed at 180°, the value of $J(^{11}B–H)$ increases slightly, about 6 Hz with respect to the value for a bent OH^- group. Such a hypothetical C_{3v} structure having three equivalent B–H bonds becomes, in fact, a transition structure, being energetically unfavorable by some 20 kcal/mol above the equilibrium geometry. Thus, in the theoretical modeling of the $(BH_3OH)^-$ species one cannot assume equivalent B–H bonds. The use of the B3LYP functional leads to an overestimation of $J(^{11}B–H)$ compared to the PW91PW91 (about 10%), which indicates that the use of the PW91PW91 functional should give J values closer to the experimental. The estimated $^1J(^{11}B–H)$ coupling constants for BH_3 (135.1 Hz) differ considerably from the values for $(BH_3OH)^-$ and, therefore, the presence of BH_3 species cannot explain the experimental spectrum. To assess the accuracy of the present estimations of ^1J coupling constants, the B_2H_6 system was studied and compared to the literature values obtained by other, more accurate, methods [33]. It is seen that the present B3LYP/aug-cc- pVTZ values differ by about 10–20% from the EOM–CCSD *ab initio* estimations. Such a moderate accuracy is probably sufficient for drawing conclusions as to the relative values of ^1J(B–H) coupling constants in structurally related species such as BH_3, BH_4^-, $(BH_3OH)^-$ and $(BH_3OH)Na$. The presented theoretical calculations support the hypothesis about the detection of the $(BH_3OH)^-$ anion as an individual species in the solvent. If the complex with solvent molecules (acetone) were formed, we would probably not obtain this degree of correspondence of experimental and theoretical values of $J(^{11}B–H)$. The theoretical results suggest that the OH^- group can be located in different positions relative to the BH_3 pyramid. It is likely that the position of the hydroxyl group is averaged in solution, which renders impossible a direct comparison of theoretical and experimental J-values.

We investigated also the hypothetical complex $(BH_3OH)^- \cdots Na^+$ and the corresponding $J(^{11}B–H)$ values. The presence of the Na^+ ion influences significantly the value of the coupling of one of the protons, yielding $J(^{11}B–H)$ of about 120 Hz, which is not observed in the experimental spectrum. This suggests that the Na^+ ion

Table 1
The ^1J(B–H) coupling constants (in Hz) calculated with the B3LYP/aug-cc-pVTZ method

Species	Geometry	Bond	$^1J(^{11}B–H)$	$^1J(^{10}B–H)$	Comments
$(BH_3OH)^-$	A	B–H3	92.2	30.9	
	A	B–H4	87.5	29.3	
	A	B–H3	84.6	28.3	PW91PW91
	A	B–H4	79.7	26.7	PW91PW91
	B	B–H3	84.2		By 0.6 kcal mol less favorable than geometry A
	B	B–H4	91.5		
	C (C_{3v})	B–H	93.7	31.4	Energetically unfavorable
$(BH_3OH)Na$		B–H3	120.7	40.4	
		B–H4	84.3	28.2	
		B–H5	80.1	27.1	
BH_4^-		B–H	85.1	28.5	
BH_3		B–H	135.1	45.2	
B_2H_6		B–Hb[a]	51.2	17.2	42.7[32]
		B–Ht	132.1	44.3	122.17[32] 133.5 (experimental value)[32]

The letters A, B and C denote the different geometries of (BH3OH)-molecule:
- A is an equilibrium geometry, where the O–H group is on the plane between B–H bonds;
- B is a transition geometry, where the O–H group lies directly above one of the B–H bonds;
- C is a geometry where the B–O–H bonds are linear and the molecule has C3v symmetry.

The numbering of hydrogen atoms is as follows:
- A: H3 is connected to B and is the farthest from the O–H group H4 is equivalent to H5, the O–H group lies above them.
- B: H3 is directly below the O–H group, H4 is equivalent to H5.

[a] Hydrogen atoms bonded to both boron atoms.

can be coordinated by the acetone molecules rather than by the $(BH_3OH)^-$ anion. One can hypothesize that the Na^+ ion is solvated strongly by the acetone molecules, which reduces the possibility of direct interaction with the $(BH_3OH)^-$ ion. The detailed structure of the acetone solution with the presence of $(BH_3OH)Na$ would, however, require much more advanced simulations.

Conclusions

The present high yield synthetic route providing hydroxyborohydride salts is fast and economic, making hydroxyborohydrides quite attractive reagents. The first ^1H NMR spectrum of hydroxyborohydride shows a very unusual pattern of resonances from heteronuclear coupling of the boron-hydride moiety. An NMR spectroscopy-monitored experiment demonstrated stronger reducing properties of hydroxyborohydride than borohydride toward acetone carbonyl. The calculated ^{11}B- and ^{10}B–H coupling constants are in a very good agreement with experimental data, and confirm the hypothesis about the detection of the $(BH_3OH)^-$ anion as an individual species in solution.

Acknowledgement

Supported by the Ministry of Education and Science, Poland, Grant No. N204 088 31/2052 2006-2009. The computational work was performed on the computers of Interdisciplinary Centre for Mathematical and Computational Modelling, Warsaw University. Helpful discussion of the text by Professor David Shugar is greatly acknowledged.

Appendix A. Supplementary materials

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.inoche.2007.05.029.

References

[1] Y. Segawa, M. Yamashita, K. Nozaki, Science 314 (2006) 113.
[2] T.B. Marder, Science 314 (2006) 69.
[3] M. Miyaura, A. Suzuki, Tetrahedron Lett. (1979) 3437.
[4] N. Miyaura, A. Suzuki, Chem. Rev. 95 (1995) 2457.
[5] C.H. Burgos, E. Canales, K. Matos, J.A. Soderquist, J. Am. Chem. Soc. 127 (2005) 8044.
[6] C. Lai, J.A. Soderquist, Org. Lett. 7 (2005) 799.
[7] E. Hernandez, J.A. Soderquist, Org. Lett. 7 (2005) 5397.
[8] E.J. Corey, C.J. Helal, Angew. Chem. Int. Ed. (1998) 1986.
[9] G. Bringmann, D. Menche, Acc. Chem. Res. 34 (2001) 615.
[10] A.O. King, E.G. Corley, R.K. Andersen, R.D. Larsen, T.R. Verhoeven, P.J. Reider, Y.B. Xiang, M. Belley, Y. Leblanc, J. Org. Chem. 58 (1993) 3731.
[11] T. Akiyama, H. Nishimoto, T. Kuwata, S. Ozaki, Bull. Chem. Soc. Jpn. 67 (1994) 180.
[12] E. Mercantoni, S. Alessandrini, M. Malavolta, G. Bartoli, M.C. Bellucci, L. Sambri, R. Dalpozzo, J. Org. Chem. 64 (1999) 1986.
[13] T. Kochi, T.P. Tang, J.A. Ellman, J. Am. Chem. Soc. 124 (2002) 6518.
[14] A.K. Mandal, Org. Lett. 4 (2002) 2043.
[15] O. Repic, K. Prasad, G.T. Lee, Org. Proc. Res. Dev. 5 (2001) 519.
[16] D.A. Evans, B. Côté, P.J. Coleman, B.T. Connell, J. Am. Chem. Soc. 125 (2003) 10893.
[17] P.L. Bounds, R.M. Pollack, Biochemistry 26 (1987) 2263.
[18] J.W. Reed, W.L. Jolly, J. Org. Chem. 42 (1977) 3963.
[19] J.A. Gardiner, J.W. Collat, J. Am. Chem. Soc. 86 (1964) 3165.
[20] J.A. Gardiner, J.W. Collat, J. Am. Chem. Soc. 87 (1965) 1692.
[21] G.M. Almy, R.B. Horsfall Jr., Phys. Rev. 51 (1937) 491.

[22] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery, Jr., T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople, Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford CT, 2004.

[23] G.A. Zhurko, D.A. Zhurko, ChemCraft'1.5. (1.4b), 2006. www.chemcraftprog.com. Ref Type: Computer Program.

[24] A.D. Becke, J. Chem. Phys. 98 (1993) 5648.

[25] P.J. Hay, W.R.J. Wadt, Chem. Phys. 82 (1985) 270.

[26] C.M. Breneman, K.B.J. Wiberg, J. Comput. Chem. 11 (1990) 361.

[27] E. Cances, B. Mennucci, J. Tomasi, J. Chem. Phys. 107 (1997) 3032.

[28] E. Cances, B. Mennucci, J. Tomasi, J. Chem. Phys. 109 (1998) 260.

[29] M. Cossi, V. Barone, B. Mennucci, J. Tomasi, J. Chem. Phys. Lett. 286 (1998) 253.

[30] P.C. Hohenberg, W. Kohn, L.J. Sham, S.B. Trickey (Eds.), vol. 21, Academic Press, San Diego, CA, 1990, p. 7.

[31] G. Schaftenaar, J.H. Noordik, J. Comput. Aided Mol. Des. 14 (2000) 123.

[32] S.A. Perera, M. Nooijen, R.J. Bartlett, J. Chem. Phys. 104 (1996) 3290 (B_2H_6 for comparison).

[33] J.E. Peralta, G.E. Scuseria, J.R. Cheeseman, M.J. Frisch, Chem. Phys. Lett. 375 (2003) 452.

Bioorganic Chemistry 38 (2010) 87–91

Contents lists available at ScienceDirect

Bioorganic Chemistry

journal homepage: www.elsevier.com/locate/bioorg

Preliminary Communication

The synthesis and NMR investigation on novel boron derivatives of stavudine

Tomasz Ruman [a,*], Karolina Długopolska [a], Agata Jurkiewicz [b], Katarzyna Rydel [a], Andrzej Leś [b,c], Wojciech Rode [a,d]

[a] Rzeszów University of Technology, Faculty of Chemistry, Department of Biochemistry and Biotechnology, 6 Powstańców Warszawy Ave., 35-959 Rzeszów, Poland
[b] University of Warsaw, Faculty of Chemistry, Quantum Chemistry Laboratory, 1 Pasteur St., 02-093 Warsaw, Poland
[c] Pharmaceutical Research Institute, 8 Rydygier St., 01-793 Warsaw, Poland
[d] Nencki Institute of Experimental Biology, 3 Pasteur St., 02-093 Warsaw, Poland

ARTICLE INFO

Article history:
Received 18 November 2009
Available online 4 January 2010

Keywords:
D4T
Boron nucleosides
HIV
NMR
Stavudine

ABSTRACT

Preparation and spectroscopic properties of novel boron-containing derivatives of anti-HIV agent stavudine are presented, The new compounds, (5′-*O*-(4,4,5,5-tetramethyl-1,3,2-dioxaboronate)-2′-3′-didehydro-2′-3′-dideoxythymidine and 5′-O-(dihydroxyboronate)-2′-3′-didehydro-2′-3′-dideoxythymidine), were prepared by direct reaction between stavudine and reagents containing B—H moieties – pinacolborane and borane–dimethylsulfide complexes, respectively. The boron coordination equilibrium of those compounds was analyzed by water titration monitored by NMR. Results of the DFT calculations and NMR experiments pointed to structural and electronic similarity of tetrahedral boron complexes to phosphate group.

1. Introduction

Stavudine (2′,3′-didehydro-3′-deoxythymidine; D4T), a synthetic thymidine nucleoside analogue in which the 3′-hydroxyl group is replaced by a double bond between the 2′- and 3′-carbon atoms of the pentose ring, is a nucleoside reverse transcriptase inhibitor (NRTI) [1–3], approved for its use in combination with other antiretroviral compounds for the treatment of HIV infections in USA and European Union. The drug leads to inhibition of viral reverse transcription, as do didanosine (ddI), lamivudine (3TC), zalcitabine (ddC) and zidovudine (ZDV), also belonging to the family of NRTI [1–5], with stavudine 5′-triphosphate causing competitive *vs.* dTTP inhibition of HIV reverse transcriptase and termination of DNA polymerization (due to lack of 3′-OH in pentose ring necessary to form 3′-5′-phosphodiester linkage). As current anti-HIV agents show several drawbacks, including short- and long-term adverse effects, mitochondrial toxicity, antagonism with other antiretroviral and development of drug resistant or multi drug resistant HIV strains [6,7], numerous new stavudine analogues or derivatives were synthesized and tested as antiretroviral compounds, with only a few showing promising activity. For example (i) 4′-thiostavudine [8] and 4′-ethynylstavudine [8,9] showed inhibitory activity against HIV-1 comparable to that of stavudine, (ii) 5′—OH group of stavudine esterified with antimicrobial building blocks of selected agents (ciprofloxacin, norfloxacin, isoniazide, pyrazinamide, piperazine and dimetylamine acetic acid) led to prodrugs with broad spectrum chemotherapeutic activity [6,10], (iii) stampidine, a phosphoramidate derivative of stavudine, was a potential anti-HIV agent, some 100-fold more active than stavudine and twice as active as zidovudine [11] and (iv) stampidine, 5′-stavudine-[*p*-bromophenyl methoxyalaninylphosphate], with bromine at the phenyl moiety accelerating the compound's hydrolysis yielding *N*-alaninyl-stavudinemonophosphate [12,13].

In search of new derivatives of the drug, synthesis of boron analogue of stavudine 5′-phosphate group was undertaken. As a result, two compounds were obtained, with boron atom easily forming complexes in which an additional donor lone-pair, containing coordinates to boron atom, yields tetracoordinates, more or less distorted tetrahedral forms mimicking phosphate group.

Additionally, the pinacolborane stavudine derivative (**2**) will possibly improve cell wall transfer properties due to low polarity of pinacolborane anchor. The transferred compound would probably hydrolyze to free stavudine therefore improving overall stavudine – cell transfer.

2. Results and discussion

2.1. Synthesis and NMR properties of **1**, **2** and **3**

The synthesis of **1** (stavudine, d4t) was based on previously published work of Prusoff's group, modified and optimized towards the purity of the final product [14].

* Corresponding author.
E-mail address: tomruman@prz.edu.pl (T. Ruman).

0045-2068/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.bioorg.2009.12.005

Although relatively easy method of synthesis of alkoxyboron compounds by the reaction of boric acid and given alcohol is known, rapid equilibration could result in the formation of bis- and trisalkoxy species. Our borane approach, though more expensive and time-consuming, gives products of higher purity. Presented method of preparation of **2** and **3** is based on reaction environment without polar and protic solvents so equilibration between alkoxy and hydroxy forms is minimized.

Compound **2**, 5′-*O*-(4,4,5,5-tetramethyl-1,3,2-dioxaboronate)-2′-3′-didehydro-2′-3′-dideoxythymidine, a 5′-pinacolborane derivative of stavudine was synthesized by direct reaction between a large excess (25-fold molar excess over **1**) of pinacolborane (4,4,5,5-tetramethyl-[1,3,2]dioxaborolane) and unprotected **1** (Fig. 1). The obtained hydrophobic boronic ester can be easily isolated from reaction mixture by a standard crystallization or extraction methods.

The reaction mechanism is typical for most of the borane complexes with B—H group and involves attack of boron-attached hydrogen atom onto 5′-hydroxyl group hydrogen atom with evolution of gaseous hydrogen, followed by an attack of resultant 5′-alcoholate oxygen onto boron atom. Compound **3** was prepared in a similar manner, with **1** being reacted with borane–dimethylsulphide complex (Fig. 1) at 0 °C, followed by hydrolysis of the resulting hydride compound 5′-CH_2—O—BH_2. The proposed reaction mechanism is similar to the one described for synthesis of **2**, with significant reactivity differences caused by lower electron density on the boron-attached hydrogen atom of pinacolborane, resulting from strong electron-withdrawing effect of the two pinacolborane oxygen atoms. It should be noted that no hydroboration reaction was observed under the reaction conditions used.

The ^{1}H NMR spectra of compounds **2** and **3** are similar to that of stavudine (**1**) with significant changes in the saccharide C(5′) region containing two magnetically nonequivalent hydrogen atoms (see Fig. 2).

Resonances of 5′-hydrogens of **1** form a multiplet at 3.67 ppm, whereas the hydrogen atoms at the corresponding position of compound **2** resonate at 4.05. In case of **3**, the steric and electronic differentiation between 5′-methylene group hydrogen atoms is large enough to form two distinct multiplets at 3.80 and 3.93 ppm. Such differentiation has not been observed for stavudine, the precursor of **3**. ^{11}B NMR spectra of **2** and **3** contain broad resonances at 22.4 and 17.8 ppm for **2** and **3**, respectively, in accord with previously published data [15–18].

2.2. DFT calculations

A possibility of tetrahedral boron group to mimic the geometric and electronic properties of phosphate group was tested with the aid of NMR and DFT methods. The close steric and electronic similarity of phosphates and esters of tetrahedral boric acid complexes (Fig. 3, bottom part of figure) might have interesting implications, such as a possibility to use boron analogs of nucleotides.

Coordination equilibrium of compounds **2** and **3** is presented in Fig. 3. Additional electron pair donors, e.g. water, thiol, amine, alcohol and also hydroxyl or carboxylate anion coordinates to positively polarized boron atom, creating a very stable eight-electron

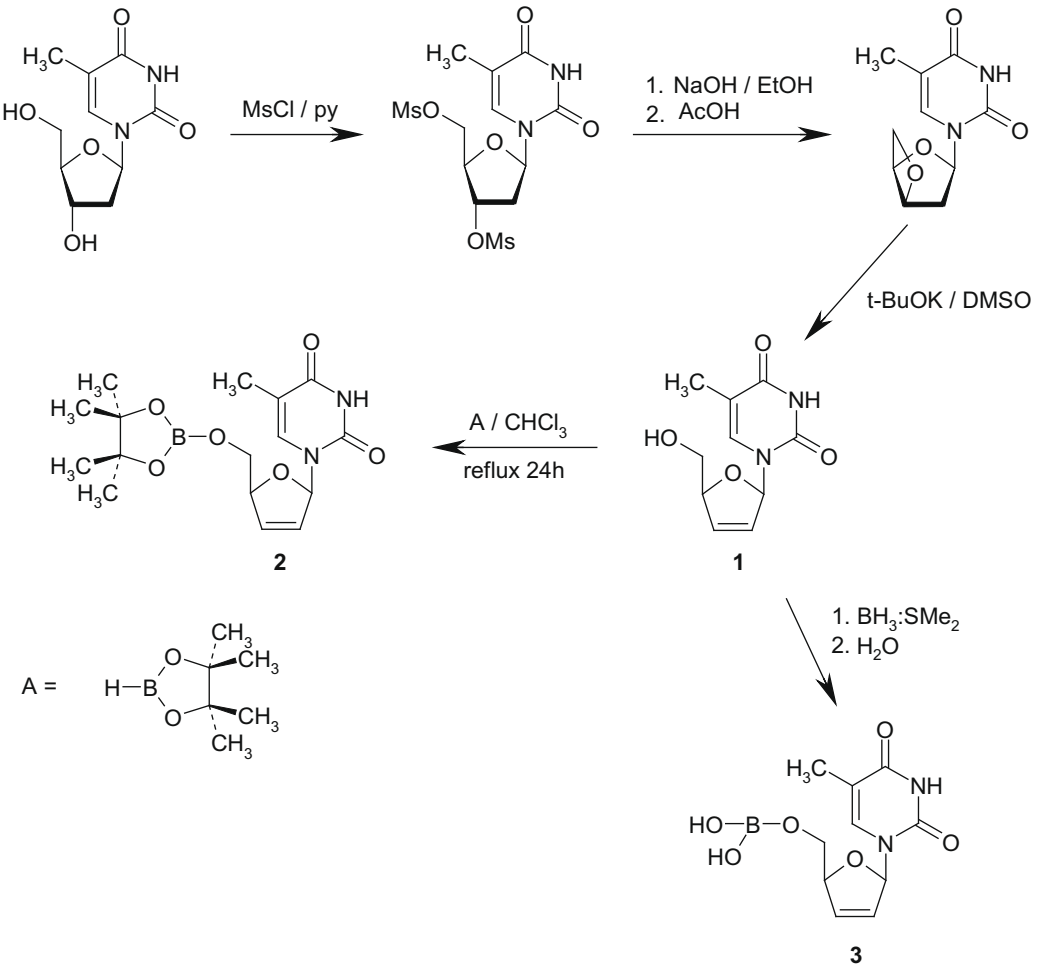

Fig. 1. Synthesis of **1**, **2** and **3**. A = 4,4,5,5-tetramethyl-[1,3,2]dioxaborolane.

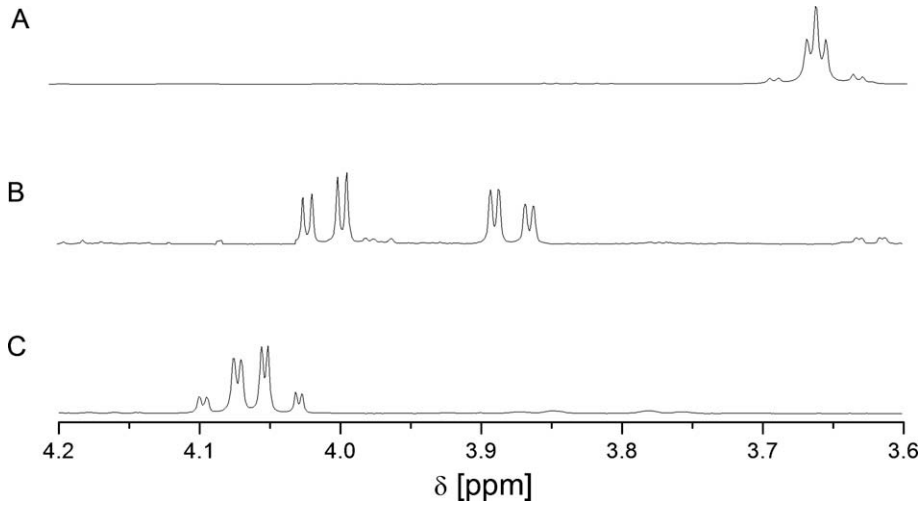

Fig. 2. ^{1}H NMR spectra of the saccharide C(5′) region of **1** (A), **2** (C) and **3** (B).

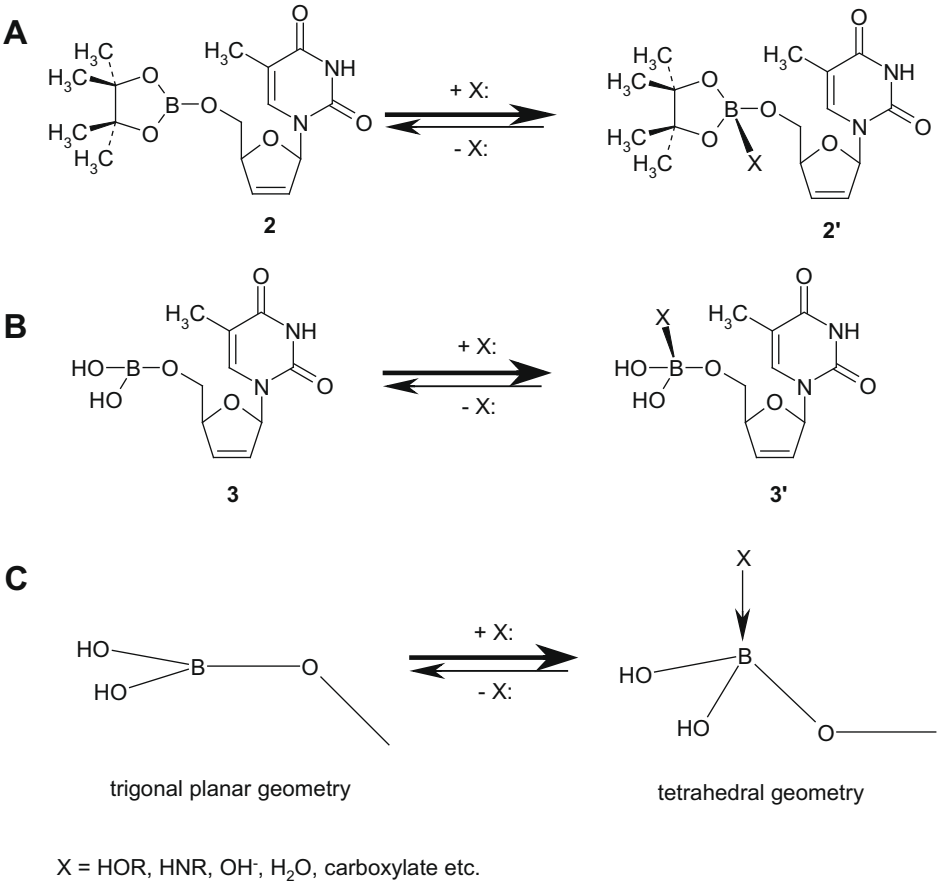

Fig. 3. Coordination equilibrium of trigonal planar and tetrahedral geometry of boron compounds with additional electron pair donor.

tetrahedral adduct. As a consequence of the high stability of eight-electron products, is the fact that vast majority of boron compounds contain only traces of trigonal form. Moreover, stable eight-electron tetrahedral complexes are often formed by intra- or extra-molecular interactions with bond lengths of 1.5–2 Å, found in X-ray diffraction data [19]. The stability of final products depends on steric and electronic properties of boron-attached groups [20].

Coordination equilibrium (*vide supra*) was confirmed with the use of the ^{11}B NMR method. The observed ^{11}B resonance of **2** was found at approx. 22.4 ppm in non-coordinating solvent. Addition of 50 μl of H_2O into NMR tube (approx. 20 mg of **2**) resulted in a significant *upfield* shift of broad ^{11}B resonance to −10.0 ppm, with additional 100 μl of H_2O shifting the described resonance even more *upfield*, up to −11.2 ppm. Titration of the sample with additional amounts of water did not result in further shifting of

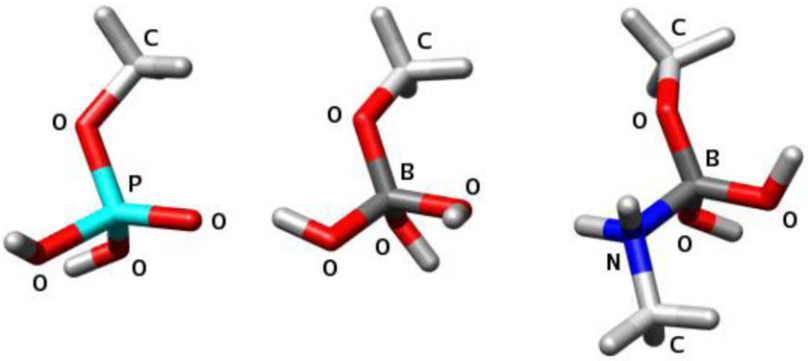

Fig. 4. The optimized geometries of stavudine-5′-CH_2—O—X models: methylphosphate (X = phosphate group, left) and its boron analogues: methylboronic acid–hydroxyl complex (X = —$B(OH)_3^-$, middle) and methylboronic acid–methylamine complex (X = —$B(OH)_2$:NH_2CH_3, right).

^{11}B resonance. Those observations confirm the equilibrium hypothesis (Fig. 3), assuming the formation of new compound (**2′**·H_2O) of tetrahedral geometry that dominates after addition of water. Similar behavior was also observed for **3**, apparently forming **3′**·H_2O. It should be noted that no significant changes were observed in the ^{1}H NMR spectrum during H_2O titration, indicating that no side reactions took place.

In accord, the ^{11}B chemical shift was previously observed to be strongly dependent on electronic environment of boron atom, with formation of tetrahedral complexes, resulting from coordination of additional lone-pair or anionic donor, shifting boron resonances *upfield* even by 20–40 ppm [15–18,20].

Considering tetrahedral boron moieties as potentially capable of mimicking the steric and also electronic properties of phosphate group, the geometries and partial atomic charges of phosphate and borane moieties were compared by theoretical calculations on three model molecules: methylphosphate, methylboronic acid–hydroxyl complex and methylboronic acid–methylamine complex (Fig. 4). The density functional method B3LYP/LANL2DZ was used to optimize geometries and to calculate Merz–Kollman ESP charges [21,22] and molecular volume.

All three molecules (Fig. 4) show tetrahedral geometry. The calculated Merz–Kollman charges for external (hydroxyl and P═O) oxygen atoms are compared in Table 1. For geometry comparisons, three parameters were used: Y—O bond lengths, O—Y—O bond angles and Y—O—C bond angles, where Y denotes P or B atom in model molecules. The bond angles are similar for all three molecules, whereas the B—O bond lengths are by about 0.2 Å smaller than P—O. Molecular volume was calculated for molecules 1 and 2. The obtained volumes were similar too, being 72.6 cm^3/mol for methylphosphate and 69.3 cm^3/mol for methylboronic acid–hydroxyl complex. The theoretical results suggest that local structure of phosphate and borane moieties should exhibit similar short-range (steric, determined by molecular volume, bond lengths and angles) as well as long-range (mainly electrostatic, determined by electronic density distribution roughly characterized by Merz–Kollman charges) intermolecular interactions with the environment. In consequence, one can expect a similar activity of phosphate and borane moieties in selected biochemical systems.

3. Conclusions

Preparation routes of novel boron-containing derivatives of anti-HIV agent stavudine (5′-*O*-(4,4,5,5-tetramethyl-1,3,2-dioxaboronate)-2′-3′-didehydro-2′-3′-dideoxythymidine and 5′-O-(dihydroxyboronate)-2′-3′-didehydro-2′-3′-dideoxythymidine) are presented. The new compounds, were prepared by one-step reaction between stavudine and commercially available pinacolborane or borane–dimethylsulfide complexes. Resonances of 5′-hydrogens of **1** form a multiplet at 3.67 ppm, whereas the hydrogen atoms at the corresponding position of compound **2** resonate at 4.05. In case of **3**, the steric and electronic differentiation between 5′-methylene group hydrogen atoms is large enough to form two distinct multiplets at 3.80 and 3.93 ppm. ^{11}B NMR spectra of **2** and **3** contain broad resonances at 22.4 and 17.8 ppm for **2** and **3**, respectively. Coordination equilibrium was analyzed with the use of the ^{11}B NMR method. The observed ^{11}B resonance of **2** was found at approx. 22.4 ppm in non-coordinating solvent. Addition of small amount of water into NMR tube with **2** or **3** solutions resulted in a significant *upfield* shift of broad ^{11}B resonances. Those observations confirm the equilibrium hypothesis assuming the formation of new compound of tetrahedral geometry that dominates after addition of water. The theoretical results suggest that local structure of phosphate and borane moieties should exhibit similar short-range as well as long-range intermolecular interactions with the environment.

4. Experimental

4.1. Analytical methods

All NMR spectra were obtained with 500 MHz Bruker Avance II spectrometer operating in the quadrature mode. The residual peaks of deuterated solvents were used as internal standards. Structures of **1**, **2** and **3** were also confirmed by several 2-dimensional spectra: ^{1}H–^{13}C Heteronuclear Single Quantum Coherence (HSQC); ^{1}H–^{1}H Correlation spectroscopy, (COSY) and Distortionless

Table 1
Selected atomic charges and geometric properties of stavudine-5′-CH_2—O—X models. Y = P or B atom.

Partial atomic charges	Methylphosphate	Methylboronic acid–hydroxyl complex	Methylboronic acid–methylamine complex
O (external O—H or ═O)	−0.9/−0.8/−0.7 (for ═O)	−0.9/−1.0/−1.0	−1.0/−0.92
Geometry			
Distances Y—O	1.7/1.7/1.69/1.58 (for P═O)	1.5/1.5/1.5/1.5	1.46/1.45/1.46/1.68 (for B—N)
Angles O—Y—O	106.2/99.5/115.2	105.5/114.4/108.13	115.1/116.2/99.9
Angle Y—O—C	120.5	117.0	121.3

Enhancement by Polarization Transfer (DEPT) experiments. All ^{11}B NMR spectra were recorded using pure quartz 5 mm NMR tube.

Elemental analysis was performed using Elementar Vario EL-3 analyzer. All other reagents and deuterated solvents of the highest commercially available grade were purchased from Aldrich and used without further purification. Rubber septa joints were also purchased from Aldrich. All procedures, including preparation of samples for the NMR measurements, were carried out under nitrogen atmosphere. All the computations were performed with the Gaussian 03 program [23] on the computer cluster of ICM UW, Warsaw.

4.2. Modified synthetic route for 2′-3′-didehydro-2′-3′-dideoxythymidine (stavudine, D4T, **1**)

Thymidine (5 g, 20.6 mmol) was added at −15 °C to solution of methylsulfonyl chloride (MsCl, 3.52 ml, 45.4 mmol) in pyridine (43 ml). After 60 min. of vigorous stirring, water (1–2 °C, 600 ml) was added to the reaction mixture. Resulting mixture was stirred at 4 °C for another 24 h, and filtered off. White product was then washed with three 10 ml rations of water (cooled to 0 °C) and dried under high vacuum.

Reaction product was dissolved in sodium hydroxide solution (400 ml of 1.06 M solution) and refluxed for 2 h. Acetone (300 ml) was added to the reaction mixture cooled to 4 °C over a 30 min period. Precipitate that formed was filtered out, washed twice with acetone (2 × 20 ml, cooled to 4 °C) and vacuum dried. The product, in the form of white/grayish crystals, was used in the following reaction step without further purification. The crystals (0.94 g) were dissolved in anhydrous DMSO (5 ml) and a solution of potassium *tert*-butoxide (1.02 g in 2 ml of DMSO) was added dropwise. After the mixture was allowed to react for 2 h at room temperature, glacial acetic acid (1:1 v/v in ethanol) was used to neutralize it to pH ≈ 7, and the product was extracted with hexane (2 × 20 ml). Hexane layer was separated and dried under high vacuum. The resulting compound **1**, obtained with 42% yield, showed 99.5% purity, as judged from the ^1H NMR spectrum. Elemental analysis: C 57.62% (calculated 57.69%) H 5.87% (calculated 5.81%); N 13.42% (calculated 13.45%). ^1H NMR ($CDCl_3$, δ [ppm]): 7.50 (m, J = 1.3 Hz, 1H, C(6)H); 6.82 (m, J = 1.6 Hz, 1H, C(1′)H); 6.35 (m, 1H, C(2′)H); 5.86 (m, 1H, C(3′)H); 4.87 (m, 1H, C(4′)H); 3.67 (m, 2H, C(5′)H); 1.74 (d, J = 0.9 Hz, 3H, Me-C(5)).

4.3. Synthesis of 5′-O-(4,4,5,5-tetramethyl-1,3,2-dioxaboronate)-2′-3′-didehydro-2′-3′-dideoxythymidine (**2**)

1 (25 mg) was dissolved in anhydrous DMSO (50 μl) and added dropwise to the solution of 4,4,5,5-tetramethyl-1,3,2-dioxaborolane (384 μl in 10 ml of anhydrous $CHCl_3$). The reaction mixture was refluxed for 24 h, followed by removal of solvents under vacuum. The resulting material was left for crystallization in tetrahydrofuran (5 ml, 72 h, 4 °C), with crystals filtered off and liquid fraction dried under high vacuum. Thick oil obtained from THF fraction was containing almost pure **2**. Compound **2** was obtained with 95% yield and showed 96% purity, as judged from the ^1H NMR spectrum. Elemental analysis: C 54.85 (calculated 54.88%); H 6.69 (calculated 6.62%); N 7.99% (calculated 8.00%). ^1H NMR ($CDCl_3$, δ [ppm]): 9.48 (s, 1H, NH); 7.54 (m, J = 1.3 Hz, 1H, C(6)H); 7.00 (m, J = 1.9 Hz, 1H, C(1′)H); 6.26 (m, 1H, C(2′)H); 5.79 (m, 1H, C(3′)H); 4.88 (m, 1H, C(4′)H); 4.05 (m, 2H, C(5′)H); 1.89 (d, J = 1 Hz, 3H, Me-C(5)); 1.20 (s, 12H, $(CH_3)_4$). ^{13}C NMR ($CDCl_3$, δ [ppm]): 164.64, 151.35 ($C_{C=O}$); 137.02 (C(6)); 133.95 (C(2′); 126.97 (C(3′)); 89.34 (C(1′)); 86.02 (C(4′)); 75.14 (O-CMe_2); 65.11 (C(5′)); 29.69 ($(CH_3)_4$); 12.09 (C(5)-Me). ^{11}B NMR ($CDCl_3$, δ [ppm]): 22.37.

4.4. Synthesis of 5′-O-(dihydroxyboronate)-2′-3′-didehydro-2′-3′-dideoxythymidine (**3**)

1 (25 mg) was dissolved in anhydrous DMSO (50 μl) and added dropwise to the solution of borane–dimethylsulfide complex (120 μl in 2 ml of anhydrous CH_2Cl_2). The reaction mixture was stirred for 24 h (0 °C), followed by removal of solvents under vacuum. Deoxygenated water (69 μl) in THF (2 ml) was added dropwise to the obtained liquid material and the mixture allowed to react for 2 h, then extracted with water/chloroform (10 + 10 ml) system. Organic solvents from chloroform phase were vacuum-evaporated and the resulting solid dried under high vacuum. Compound **3** was obtained with 95% yield and showed 96% purity, as judged from the ^1H NMR spectrum in. Elemental analysis: C 44.75 (calculated 44.81%); H 4.92% (calculated 4.89%); N 10.40% (calculated 10.45%). ^1H NMR ($CDCl_3$, δ [ppm]): 8.33 (s, 1H, NH); 7.47 (m, J = 1.3 Hz, 1H, C(6)H); 7.01 (m, J = 1.9 Hz, 1H, C(1′)H); 6.32 (m, 1H, C(2′)H); 5.84 (m, 1H, C(3′)H); 4.91 (m, 1H, C(4′)H); 3.93 + 3.80 (m, 2H, C(5′)H); 1.86 (d, J = 1 Hz, 3H, Me-C(5)). ^{13}C NMR ($CDCl_3$, δ [ppm]): 163.62 ($C_{C=O}$); 136.58 (C(6)); 134.50 (C(2′); 125.53 (C(3′)); 89.92 (C(1′)); 87.08 (C(4′)); 63.42 (C(5′)); 12.43 (C(5)-Me). ^{11}B NMR ($CDCl_3$, δ [ppm]): 17.76.

Acknowledgments

Supported by the Ministry of Education and Science, Poland, Grant No. N204 088 31/2052 2006–2009. The ICM computer center, University of Warsaw, Poland, is acknowledged for the G18-6 computer grant.

References

[1] R.M.W. Hoetelmans, AIDS Rev. 1 (1999) 167–178.
[2] T.S. Skinner-Adams, J.S. McCarthy, D.L. Gardiner, K.T. Andrews, Trends Parasitol. 24 (6) (2008) 264–271.
[3] S. Orsega, J. Nurse Pract. 10 (2007).
[4] J. Balzarini, P. Heredewijn, E. De Clercq, J. Biol. Chem. 264 (1989) 6127–6133.
[5] Fatih M. Uckun, Sharon Pendergrass, Sanjive Qazi, P. Samuel, T.K. Venkatachalam, Eur. J. Med. Chem. 39 (2004) 225–234.
[6] M.J. Otto, Curr. Opin. Pharmacol. 4 (2004) 431–436.
[7] B.A. Styrt, T.D. Pizza-Hepp, G.K. Chikami, Antivir. Res. 31 (1996) 121–135.
[8] H. Kumamoto, T. Nakai, K. Haraguchi, K.T. Nakamura, H. Tanaka, M. Baba, Y.-C. Cheng, J. Med. Chem. 49 (2006) 7861–7867.
[9] K. Haraguchi, M. Sumino, H. Tanaka, J. Org. Chem. 71 (2006) 4433–4438.
[10] D. Sriram, P. Yogeeswari, N. Srichakravarthy, T.R. Bal, Bioorg. Med. Chem. Lett. 14 (2004) 1085–1087.
[11] T.K. Venkatachalam, M. Sarquis, S. Qazi, F.M. Uckun, Bioorg. Med. Chem. 14 (2006) 6420–6433.
[12] R. Vig, T.K. Venkatachalam, F.M. Uckun, Antivir. Chem. Chemother. 9 (1998) 445–447.
[13] T.K. Venkatachalam, H.L. Tai, R. Vig, C.L. Chen, S. Jan, F.M. Uckun, Bioorg. Med. Chem. Lett. 8 (1998) 3121–3125.
[14] T.-S. Lin, R.F. Schinazi, W.H. Prusoff, Biochem. Pharmacol. 36 (17) (1987) 2713–2718.
[15] G.R. Eaton, J. Chem. Educ. 46 (1969) 547.
[16] W.G. Henderson, E.F. Mooney, Annu. Rev. NMR Spectrosc. 2 (1969) 219.
[17] W.L. Smith, J. Chem. Educ. 54 (1977) 469.
[18] B. Wrackmeyer, Annu. Rep. NMR Spectrosc. 20 (1988) 61.
[19] A. Flores-Parra, R. Contreras, Coordin. Chem. Rev. 196 (2000) 85–124.
[20] D.G. Hall (Ed.), Boronic Acids, Wiley-VCH Verlag GmbH & Co., KGaA, Weinheim, 2005.
[21] B.H. Besler, K.M. Merz Jr., P.A. Kollman, J. Comp. Chem. 11 (1990) 431.
[22] U.C. Singh, P.A. Kollman, J. Comp. Chem. 5 (1984) 129.
[23] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery Jr., T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople, Gaussian 03, Revision E.01, Gaussian, Inc., Wallingford CT, 2004.

The Synthesis, Reactivity and NMR Investigation on ^{15}N-Thiophosphoramidates

Tomasz Ruman*[,a], Karolina Długopolska[a], Agata Jurkiewicz[b], Dagmara Kramarz[a], Tomasz Frączyk[c], Andrzej Leś[b,d] and Wojciech Rode[a,c]

[a]*Rzeszów University of Technology, Faculty of Chemistry, Department of Biochemistry and Biotechnology, 6 Powstańców Warszawy Ave. 35-959 Rzeszów, Poland*

[b]*University of Warsaw, Faculty of Chemistry, Quantum Chemistry Laboratory, 1 Pasteur St., 02-093 Warsaw, Poland*

[c]*Nencki Institute of Experimental Biology, 3 Pasteur St., 02-093 Warsaw, Poland*

[d]*Pharmaceutical Research Institute, 8 Rydygier St., 01-793 Warsaw, Poland*

Received February 28, 2009: Revised October 19, 2009: Accepted October 19, 2009

Abstract: Novel ^{15}N-isotope enriched potassium and diammonium thiophosphoramidates were synthesized and their spectroscopic properties, along with reactivity towards several compounds, including histidine, thymidine, glucose and 2-deoxyribose are presented. The application of quantum mechanical DFT calculations for estimation of ^{31}P NMR chemical shifts for several thiophosphoramidate ions and its derivatives are also discussed.

Keywords: NMR, thiophosphate, tiophosphoramidate, thiophosphorylation.

Phosphates are ubiquitous in biochemistry, physico-chemical properties allowing them to play a dominating role in the living world [1]. Studies on phosphorylation/ dephosphorylation and biological activity of phosphorylated compounds, as well as those aimed at drug development, often take advantage of thiophosphate (or phosphorothioate) analogues [2-5], offering, as compared to the phosphate congeners, potential phosphorus-centered chirality and increased stability towards chemical and enzymatic hydrolysis. [6-8]. This hydrolytic stability is of particular interest with studies on phosphorylation of proteins on basic amino-acids, as due to their acid-labile character phosphoramidates within proteins easily escape detection by analytical methods [9, 10]. Moreover, phosphohistidine, representing probably a major protein modification in eukaryotes [9], applied as a hapten has proved too unstable to generate antibodies [11]. Therefore methods are sought allowing preparation of thiophos-phoramidate-modified biomolecules under conditions mild enough to preserve biological function [12, 13], in hope that thiophospho-ramidates will be useful as more stable substitutes and probes for the biological function of the corresponding phosphoramidates.

In the present study we synthesized ^{15}N-thiophos-phoramidates and tested their spectroscopic properties, as well as potential to thiophosphorylate histidine, thymidine, glucose and 2-deoxyribose.

*Address correspondence to the author at the Rzeszów University of Technology, Faculty of Chemistry, Department of Biochemistry and Biotechnology, 6 Powstańców Warszawy Ave. 35-959 Rzeszów, Poland, E-mail: tomruman@prz.edu.pl

The most common method of preparing *O*- and *N*-thiophosphorylated compounds is the reaction of peptide with a suitable kinase and γ-S-ATP [17]. A few non-enzymatic approaches have been also described, namely the on-resin phosphitylation-oxidation of a protected peptide and the integration of a protected thiophosphorylated compound by traditional Boc or Fmoc synthesis methods into a peptide chain [18-20]. However, considering possible protein thiophosphorylation, aimed at monitoring of the modification's influence on biological properties, the latter methods do not allow reaction under mild enough conditions. Another approach employed $PSCl_3$ as a thiophosphorylation agent [13] and allowed peptide modification under mild conditions (room temperature, pH 8) but the hydrophobic character and wide range of reactivity of thiophosphoryl chloride may cause the reaction sites to be difficult to control. Moreover, $PSCl_3$ and its hydrolysis product $PSCl_2O^-$ could react as a cross-linking or bridging reagents. In contrast, the unifunctional reagents, anionic thiophosphoramidates, in analogy to the parent phosphoramidates [21], allow mild reaction conditions [12] and should provide greater specificity.

The synthetic route applied to obtain ^{15}N-enriched potassium thiophosphoramidate (**2**, see Fig. (**1**)) was based on that previously employed to synthesize the same compound containing natural nitrogen isotope [14]. The first reaction of the two-step synthesis is that of phosphorus thiochloride with ^{15}N-ammonium hydroxide solution yielding **1** (^{15}N-diammonium thiophosphoramidate, $(^{15}NH_4)_2PS(^{15}NH_2)O_2$). The ^{31}P NMR spectrum of **1** in D_2O presents doublet resonance at 40.4 ppm belonging to phosphorus from **1** and much smaller resonance of hydrolysis product of **1** at 32.8 ppm. The resonance of **1** is in a form of a doublet resulting from the heteronuclear coupling

1570-1786/09 $55.00+.00

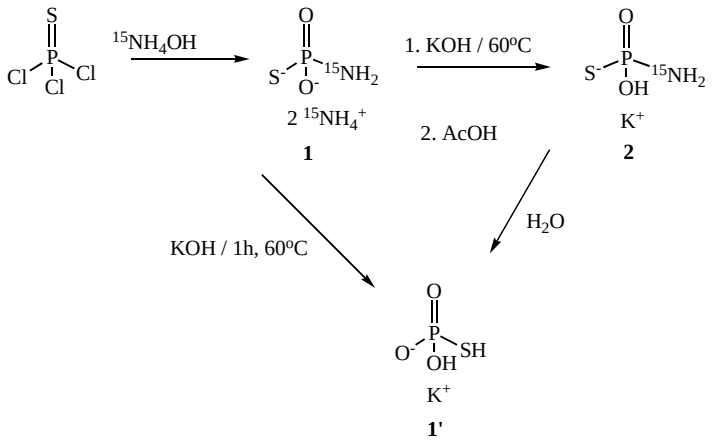

Fig. (1). Synthesis of **1**, **1'** and **2**.

between ^{31}P and ^{15}N nuclei, each having nuclear spin of 1/2. The analogical doublet resonance, corresponding to nitrogen atom of amidate moiety was found in ^{15}N NMR spectrum of **1** at 68.3 ppm. The one-bond ^{31}P-^{15}N coupling constant is 13.7 Hz being identical in both phosphorus and nitrogen spectra. Besides, the same ^{15}N NMR spectrum of **1** shows a singlet resonance of ammonium ions at 19.9 ppm.

The second step of synthesis of **2**, basically an ion exchange reaction, is conducted at temperature of 60ºC. In view of our results, this step could be potentially destructive and was probably responsible for previously reported unsuccessful synthesis of **2** [13]. The reaction with potassium hydroxide, running under those conditions longer than 5 min, leads to decomposition of thiophosphoramidate due to hydrolysis of the P-N bond. The reaction time exceeding 30min results in almost quantitative hydrolysis of thiophosphoramidate to thiophosphate (**1'**, Fig. (**1**)), the latter showing its ^{31}P NMR singlet resonance at 36.7 ppm. No nitrogen resonances were found in ^{15}N NMR spectrum of the hydrolysis product, and this observation was confirmed by other analytical methods demonstrating the absence of nitrogen in the sample. With the reaction time of approx. 5 min, the second step of synthesis of **2** gave the expected product with a good yield. The ^{31}P NMR spectrum of **2** is similar to that of **1** and presents a doublet resonance at 39.8 ppm, this chemical shift being in agreement with earlier published data for the natural-nitrogen compound [14].

Thiophosphorylation of histidine using thiophosphoramidate (Fig. (**2**)) with natural nitrogen isotopes abundance was presented by Pirrung *et al.* [12]. The histidine derivatization presented in our work was conducted with the use of ^{15}N-enriched thiophosphoramidate that gives more interesting mechanistic information. The mechanism of this reaction contains an attack of the amino acid nitrogen lone pair electrons on the phosphorus atom with elimination of the amidate group that is ^{15}N-enriched in our experiments. The ^{31}P NMR spectrum of the product shows a singlet resonance of 3-thiophosphohistidine, confirming the anticipated mechanism (*vide supra*). In accord with the latter, no ^{15}N resonances were found.

Potassium thiophosphoramidate was used for phosphorylation of saccharide hydroxyl groups of thymidine, glucose and 2-deoxyribose. Thiophosphorylation of thymidine leads to 5'-thiophosphate derivative (Fig. (**2**)) as judged from NMR spectra. The ^{31}P NMR spectrum of the reaction mixture contains peaks (43.0 ppm), in the forms of a doublet and asymmetrical singlet, which result from the three-bond heteronuclear coupling between phosphorus and two magnetically non-equivalent 5'-hydrogen atoms. The explanation of this observation is the magnetic nonequivalence of two hydrogen atoms from 5'-methylene group. The nonequivalent hydrogen atoms couple to thiophosphate phosphorus atom with different coupling constant that results in doublet and asymmetrical singlet resonances. The doublet form is a result of coupling of phosphorus of ½ spin with one hydrogen atom of ½ spin. Usually for saccharide methylene group one observes two $^{3}J_{P\text{-}H}$ couplings with different constants. The larger coupling constant might result in clearly doublet form resonance, while smaller coupling constant may give doublet or unsymmetrical singlet resonance depending on resolution of the spectrum.

The $^{3}J_{P\text{-}H}$ coupling constant for doublet resonance is in the typical range for heteronuclear couplings of this type - 11.3 Hz [22]. The ^{31}P NMR spectrum with hydrogen-decoupling applied shows only one singlet resonance pointing to the observed resonance pattern to originate from a long-range P-H coupling. The latter observation was confirmed by application of the gradient-enhanced ^{1}H-^{31}P heteronuclear multiple bond correlation (HMBC) experiment showing a strong coupling with 5'-hydrogen atoms at 3.82 ppm, in accord with previously published data [23].

Similar thiophosphorylation reaction was conducted for glucose (Fig. (**2**)). The obtained ^{31}P spectra present mainly two doublets (40.0 and 39.4 ppm) with $^{3}J_{P\text{-}H}$ values of 10.3 Hz and 10.4 Hz for glucose-1-thiophosphate α- and β-anomers, respectively. Similarly to thiophosphorylated thymidine, the hydrogen decoupling reduces two doublets to two singlets at averaged chemical shift positions proving that the doublet resonance forms are due to proximity of 1'-hydrogen atom. The gradient-enhanced ^{1}H-^{31}P HMBC experiments allowed to find chemical shifts of hydrogen atoms participating in coupling with phosphorus. The α- and β-anomers of glucose-1-thiophosphate are in 1 to 9.5 molar

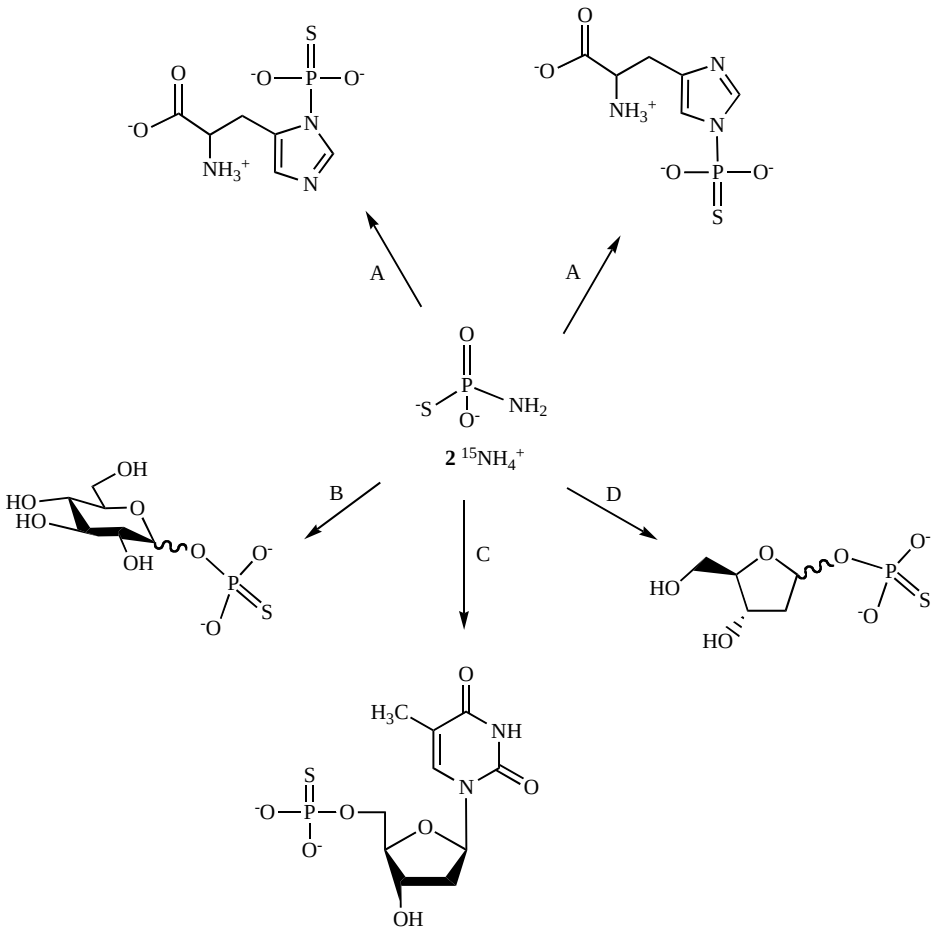

Fig. (2). Thiophosphorylation products obtained by reaction with thiophosphoramidates. **A** = histidine, **B** = D-(+)-glucose, **C** = thymidine, **D** = 2-deoxy-D-ribose.

ratios respectively as calculated from integrals of resonances [24, 25].

The ^{31}P NMR spectrum of 2-deoxyribose thiophosphorylation product (Fig. (**2**)) also contains two doublets at 38.8 and 38.9 ppm with $^{3}J_{P\text{-}H}$ values of 11.3 and 10.4 Hz for 2-deoxyribose-1-thiophosphate α- and β-anomers, respectively. The α- and β-anomers of glucose-1-thiophosphate are in 1 to 2.5 molar ratios respectively. Similarly to glucose thiophosphates, the gradient-enhanced ^{1}H-^{31}P HMBC and hydrogen-decoupled ^{31}P 1D experiment allowed the analysis of the composition of the reaction mixture [25, 26].

In view of the foregoing, as well as previously published data, potassium and ammonium thiophosphates are derivation agents with a broad spectrum of use. Of particular importance is fact that both compounds may be used under conditions that should preserve protein structure. Recently we found both potassium and diammonium thiophosphoramidates capable to thiophosphorylate His, Ser and Lys residues (not shown) of *Caenorhabditis elegans* recombinant thymidylate synthase protein [27].

Several model systems have been chosen for comparison and for assessment of trends rather than attempting to reproduce absolute values of NMR parameters. Such a strategy stems from the fact that the calculation of shielding constants and spin-spin coupling constants is extremely sensitive to the choice of the molecular basis sets.

For most of the theoretical calculations the B3LYP/aug-cc-pVTZ density functional theory was used. Due to the lack of the standard aug-cc-pVTZ basis set for potassium, the Sadlej's pVTZ basis set was used instead [28-32]. During the course of geometry optimizations it was noted that the mercaptophosphinyl tautomeric form of molecule **1'** (containing the P-SH bond) is less stable by 43.6 kJ mol^{-1} (10.4 kcal mol^{-1}) than the phosphinothioyl form, containing the formal P=S group. The potassium ion in this molecule is located between the two nucleophilic centres: the O atom and S atom (*Supplementary Materials*, Fig. (**1A**)). Another interesting case was found for molecule **1** where the protons originally attached to the NH_4^+ cations were finally transferred to the oxygen atoms and the complex of thiophosphoamide with two NH_3 molecules arised (*Supplementary Materials*, Fig. (**2A**)). The calculated isotropic shielding constants are included in Table **1**. The degree of ionization of molecule **1** in the course of hydrolysis is expected to be reflected in values of the ^{31}P isotropic shielding constants. For comparison, the ^{31}P isotropic shielding constants, calculated with the B3LYP/aug-cc-pVTZ method, for phosphoric acid and its ions are: 310.73 for H_3PO_4; 304.18 for $H_2PO_4^-$, 286.57 for HPO_4^{2-} and 272.76 for PO_4^{3-}. The isotropic shielding

Table 1. Isotropic Shielding Constants [ppm] Estimated with the B3LYP/aug-cc-pVTZ GIAO Calculations

	1			**2**		**1'**			
						-OH, =S		**=O, -SH**	
	Q = 0	**Q=-1**	**Q = -2**	**Q = 0**	**Q = -1**	**Q = 0**	**Q = -1**	**Q = 0**	**Q = -1**
Atom									
P	244.3	258.0	266.7	245.8	247.0	247.1	245.7	297.9	295.3
S	648.3	617.7	643.6	624.4	641.7	649.0	654.2	553.6	559.6
O	170.7	168.0	126.6	130.5	146.8	137.0	149.1	111.0	108.3
O	163.1	147.5	126.6	184.4	173.2	169.8	145.4	118.1	133.8
O	-	-	-	-	-	188.1	176.5	169.5	164.8
N	187.0	170.1	143.2	176.7	164.4	-	-	-	-
δ_P [ppm] rel. to H_3PO_4	66.4	52.7	44.0	64.0	63.7	63.6	65.1	12.8	15.4
δ_P [ppm] rel. to HPO_4^{2-}	42.3	28.5	19.9	40.8	39.5	39.5	40.9	11.4	8.7

The molecular total electric charge Q [a.u.] corresponds to the molecular complexes shown in the Scheme **1** (Q = 0) and to their respective anions (Q = -1 or -2) after removal of counter ions (NH_4^+ or K^+). For **1'** two geometries were considered where the P=S and P-SH bonds can appear. δ_P stands for ^{31}P chemical shift in relation to H_3PO_4 or HPO_4^{2-}.

constants calculated for $PSCl_3$ are 237.61 ppm (P) and 280.22 ppm (S). For molecule **2** the spin-spin coupling constant was also calculated, giving $^1J_{P\text{-}N}$=17.7 Hz.

As the accurate and reliable estimation of chemical shifts for ^{31}P NMR method is still long-term challenge, there is very little knowledge describing thiophosphates and their derivatives. The estimation of phosphorus chemical shift δ_P (Table **1**, lowest line) based on comparison of electronic shielding of particular atoms can be very useful information regarding NMR properties of thiophosphates, a subject poorly described in literature. The calculated ^{31}P chemical shifts for **1**, **2** and **1'** in relation to H_3PO_4 (0.0 ppm) are having similar values lying far above 35-45 ppm range which is typical for thiophosphates. The only one value of calculated chemical shift for **1**-dianion seems to be in thiophosphate range mentioned earlier. The similar comparison of shieldings in relation to HPO_4^{2-} dianion gives much better results. The calculated chemical shifts for **1**, **2** and **1'** (Q=0) in this case are 42.3, 40.9 and 39.5 ppm respectively and are in much better agreement with experimental results. Another important observation can be made by comparison of shieldings of compounds with formal phosphorus-sulfur double and single bonds. The electronic shielding of ion **1'** has been calculated with both formal double P=S bond with protonated oxygens (phosphinothioyl tautomeric form, columns 6 and 7 in Table **1**) and with single P-S bond (mercaptophosphinyl tautomeric forms, columns 8 and 9 in Table **1**). Observed and calculated differences between those two forms are striking. The calculated chemical shift of phosphinothioyl tautomer lays in range typical for thiophosphates, but the P-SH tautomeric form is *upfield* shifted and lies in 8-18 ppm range. This observation can be related to very similar NMR properties of S-substituted thiophosphate ions, for example with S-thiophosphorylated sugars [25]. In conclusion, DFT methods can be very useful anticipating NMR chemical shifts of newly synthesized thiophosphate compounds as their ^{31}P NMR chemical shifts are found in quite large range starting from +5 ppm, even to +80 ppm.

The importance of thiophosphate modification of proteins and also smaller biologically active compounds has been presented several times in literature. This paper demonstrates that thiophosphoramidates are very promising and useful derivation agents with vast range of possible application. We have presented synthetic route to obtain novel ^{15}N-enriched potassium thiophosphoramidate and its NMR properties with the use of both ^{31}P- and ^{15}N NMR methods. The published synthetic route for potassium thiophosphoramidate has been evaluated and optimized. It was shown with the use of ^{15}N NMR method that the mechanism of thiophosphorylation of histidine and presumably other amino acids involves an attack of the amino acid nitrogen lone pair electrons on the phosphorus atom with the elimination of amidate group. There are also presented NMR properties of *O*-thiophosphorylation products of thymidine, glucose and 2-deoxyribose. We also present series of calculation results with the aim to estimate the ^{31}P NMR chemical shifts by analyzing electronic shieldings of thiophosphoramidate and thiophosphate ions. It has been shown that the chemical shift of thiophosphate ions in phosphinothioyl tautomeric form lays in a range typical for thiophosphates but the mercaptophosphinyl tautomer is *upfield* shifted and lays in 8-18 ppm range, typical for S-substituted thiophosphates.

EXPERIMENTAL SECTION

All NMR spectra were obtained with Bruker Avance spectrometer operating in the quadrature mode at 500.13 MHz for ^1H, 202.46 MHz for ^{31}P and 50.70 MHz for ^{15}N nuclei. The residual peaks of deuterated solvents were used as internal standards in ^1H NMR method. ^{31}P and ^{15}N NMR spectra were recorded at 298K both with and without proton decoupling. The internal standard used in *^{31}P NMR* was inorganic phosphate (P_i) having its resonance at 2.14-2.16 ppm (at pH 7.8), 2.00-2.05 ppm (pH 7.5), 1.60-1.70 ppm (at pH 5.0) and 0.0 ppm (at pH 1-1.5). All samples were analyzed using gradient-enhanced ^1H-^{31}P Heteronuclear Multiple Bond Correlation (HMBC) experiments. The HMBC experiments were optimized for long range

couplings by using various $^3J_{P-H}$ values (1-20Hz). The ^1H NMR spectra were obtained with the use of HDO suppression method. Elemental analysis was performed using Elementar Vario EL-3 analyzer. FTIR spectra were recorded on Perkin Elmer Paragon 1000 apparatus. The ^{15}N-ammonium hydroxide solution (^{15}N>99%) was purchased from SpectraGases Inc. Potassium thiophosphoramidate was obtained using previously described method, slightly modified with regard to the time of reaction (5-15 min) with potassium hydroxide [14]. All other reagents and deuterated solvents of the highest commercially available grade were purchased from Aldrich and used without further purification. Rubber septa joints were also purchased from Aldrich. All procedures, including preparation of samples for the NMR measurements, were carried out under nitrogen atmosphere. The theoretical calculations have been performed with the density functional B3LYP/aug-cc-pVTZ method. The optimal geometries were obtained and confirmed with positive harmonic frequencies. All the calculations were performed with the Gaussian G03 (rev. C.02) suite of programs [15]. The pictures of calculated structures were made in the ChemCraft computer program and are presented in *Supplementary Materials* [16].

^{15}N-Thiophosphoramidate Diammonium Salt (1)

Thiophosphoryl chloride (0.25 ml, 2.45 mmol) was added at 0ºC to 3N solution of $^{15}NH_4OH$ (4 ml, 12 mmol). After 90 min. of vigorous stirring, the reaction mixture was separated in a separatory funnel and to the aqueous layer acetone (7 ml) added. Resulting mixture was stirred at 0ºC for 60 min and filtered, to collect white precipitate that formed. The product was then washed with two 10 ml rations of tetrahydrofurane (cooled to 0ºC), then two 10 ml rations of ethyl ether, and dried under high vacuum. Yield: 0.212 g (58%). Elemental analysis: H 6.68% (calculated 6.71%); N 29.92 (calculated 29.98%). ^{31}P NMR (D_2O, δ [ppm]): 38.02 (**1**, d, $^1J_{31P-15N}$ = 13.7 Hz); 53.60 ($PSCl_3$, s); ^{15}N NMR (D_2O, δ [ppm]): 68.34 (**1**, d, $^1J_{15N-31P}$ = 13.7 Hz); 19.92 ($^{15}NH_4^+$, s, LW 0.9 Hz).

^{15}N-Potassium Thiophosphoramidate (2)

1 (0.2 g) was dissolved in 5 ml of 10% KOH and stirred at 60ºC for 5 min. The solution was then cooled to room temperature and brought to pH 6 with acetic acid. Ethanol (20 ml) was added (at 0ºC), the solution stirred for 30 min, and the resulting white precipitate collected by filtration, washed twice with tetrahydrofurane (10 ml rations) and dried under high vacuum. Elemental analysis: H 1.97% (calculated 1.99%); N 9.84% (calculated 9.86%). ^{31}P NMR (D_2O, δ [ppm]): 39.79 (d); FTIR (KBr, cm^{-1}): 3391.8, 1656.3, 1571.9, 1410.2, 1091.0, 898.2, 871.8, 650.3, 623.3, 536.9.

Hydrolysis of ^{15}N-Potassium Thiophosphoramidate

2 (50 mg) was dissolved in 0.6 ml of 50% KOH and stirred at 60ºC for 60min. The solution was then cooled to room temperature and brought to pH 6 with acetic acid. Ethanol (20 ml) was added to obtained solution (at 0ºC), and stirred for 30min. The white precipitate was collected by filtration, washed with tetrahydrofurane (2 times 10 ml) and dried under high vacuum. Potassium thiophosphate (KH_2PSO_3, **1'**) was obtained as a main product. Elemental analysis: H 1.40% (calculated 1.32%); N 0.0% (calculated 0.0%). ^{15}N NMR (D_2O, δ [ppm]): no ^{15}N resonances found.

^{31}P NMR (D_2O, δ [ppm]): 36.70 (s).

^{15}N-Thiophosphorylation of Histidine

To a sample of 10 mg histidine in 0.2M Tris-HCl buffer (pH 7.5, 200 μL) potassium ^{15}N-thiophosphoramidate was added (2:1 potassium thiophosphoramidate to amino acid molar ratio) and the reaction mixture shaken for 24 h at 298K. The NMR sample contained 200μL of reaction solution and 400 μL of D_2O. ^{31}P NMR (D_2O, δ [ppm]): 34.84 (s).

Thiophosphorylation of α-D-Glucose

To a sample of 100 mg glucose in 300 μL of D_2O potassium thiophosphoramidate was added (3:1 potassium thiophosphoramidate to saccharide molar ratio) and the reaction mixture shaken for 24 h at 298K. The NMR sample contained the reaction mixture and 300 μL of D_2O. ^{31}P NMR (D_2O, δ [ppm]): 39.35 (d, J=10.4Hz); 39.96 (d, J=10.3Hz).

Thiophosphorylation of Thymidine

To the sample containing 100 mg thymidine in 300 μL of D_2O, the potassium thiophosphoramidate was added (2.3:1 potassium thiophosphoramidate to saccharide molar ratio). The sample was then shaken for 24 hours at 298K. The NMR sample contained reaction mixture and 300 μL of D_2O. ^{31}P NMR (D_2O, δ [ppm]): 43.00 (d+s, J_d=11.3Hz).

Thiophosphorylation of 2-Deoxyribose

To a sample of 100 mg 2-deoxyribose in 300 μL of D_2O potassium thiophosphoramidate was added (2.1:1 potassium thiophosphoramidate to saccharide molar ratio) and the resulting mixture shaken for 24 h at 298K. The NMR sample contained the reaction mixture and 300 μL of D_2O. ^{31}P NMR (D_2O, δ [ppm]): 38.75 (d, J=11.3Hz); 38.93 (d, J=10.4Hz).

ACKNOWLEDGEMENTS

Supported by the Ministry of Education and Science, Poland, Grant No. N204 088 31/2052 2006-2009. The ICM computer center, University of Warsaw, Poland, is acknowledged for the G18-6 computer grant.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publishers Web site along with the published article.

REFERENCES

[1] Westheimer, F.H. *Science*, **1987**, *235*, 1173.
[2] Eckstein, F. *Annu. Rev. Biochem*., **1985**, *54*, 367.
[3] Eckstein, F.; Gish, G. *Trends Biochem. Sci.*, **1989**, *14*, 97.
[4] Heidenreich, O.; Pieken, W.; Eckstein, F. *FASEB J.*, **1993**, *7*, 90.

[5] Mescalchin, A.; Detzer, A.; Wecke, M.; Overhoff, M.; Wünsche, W.; Sczakiel, G. *Expert Opin. Biol. Ther.*, **2007**, *7*, 1531.
[6] Woźniak, L.A.; Okruszek, A. *Chem. Soc. Rev.*, **2003**, *32*, 158.
[7] Allen, J.J.; Lazerwith, S. E.; Shokat, K. M. *J. Am. Chem. Soc*., **2005**, *127(15)*, 5288.
[8] Zhao, Z. *BBRC*, **1996**, *218*, 480.
[9] Matthews, H.R. *Pharmacol. Ther*., **1995**, *67*, 323.
[10] Klumpp, S.; Krieglstein, J. *Biochim. Biophys. Acta*, **2005**, *17*, 291.
[11] Hunter, T. in *Life Sciences for the 21st Century;* E. Keinan, I. Schechter, M. Sela (eds.), Wiley-VCH; Weinheim, **2004**; pp. 191-224.
[12] Pirrung, M.C.; James, K.D.; Rana, V.S. *J. Org. Chem*., **2000**, *65*, 8448.
[13] Lasker, M.; Bui, C.D.; Besant, P.G.; Sugawara, K.; Thai, P.; Medzihradszky, G.; Turck, C.W. *Protein Sci.*, **1999**, *8*(10), 2177.
[14] Katagi, T. *J. Mol. Struct.: Theochem*, **1990**, *68*, 61.
[15] Gaussian 03, Revision C.02; Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr. J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S.S.; Tomasi, J. ; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M. ; Toyota, K.; Fukuda, R. ; Hasegawa, J.; Ishida, M. Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C. ; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A.J.; Cammi, R.; Pomelli, C.; Ochterski, J.W.; Ayala, P.Y.; Morokuma, K.; Voth, G.A.; Salvador, P.; Dannenberg, J.J.; Zakrzewski, V.G.; Dapprich, S.; Daniels, A.D.; Strain, M.C.; Farkas, O.; Malick, D.K.; Rabuck, A.D.; Raghavachari, K.; Foresman, J.B.; Ortiz, J.V.; Cui, Q.; Baboul, A.G.; Clifford, S.; Cioslowski, J.; Stefanov, B.B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R.L.; Fox, D.J.; Keith, T.; Al-Laham, M.A.; Peng, C.Y.; Nanayakkara, A.; Challacombe, M.; Gill, P.M.W.; Johnson, B.; Chen, W.; Wong, M.W.; Gonzalez, C.; Pople, J.A., Gaussian, Inc., Wallingford CT, **2004**.
[16] http://www.chemcraftprog.com/ (accessed January 5th 2009, 9:00).
[17] Frey, P. A. *Adv. Enzymol*., **1989**, *62*, 119.
[18] Mora, N.; Lacombe, J.M.; Pavia, A.A. *Int. J. Pept. Prot. Res*., **1995**, *45*, 53.
[19] Aemissegger, A.; Carrigan, C.N.; Imperiali, B. *Tetrahedron*, **2007**, *63*, 6185.
[20] Guzacev, A. P.; Manoharan, M. *Tetrah. Lett*., **2001**, *42*, 4769.
[21] Medzihradszky, K.F.; Phillips, N.J.; Senderowicz, L.; Wang, P.; Turck, C.W. *Protein Sci.*, **1997**, *6*, 1405.
[22] Isab, A.A.; Hussain, M.S.; Akhtar, M.N.; Wazeer, M.I.M.; Al-Arfaj, A.R. *Polyhedron*, **1999**, *18*, 1401.
[23] Jankowska, J.; Sobkowska, A.; Cieślak, J.; Sobkowski, M.; Kraszewski, A.; Skawiński, J.; Shugar, D. *J. Org. Chem*., **1998**, *63*, 8150.
[24] Singh A.N.; Newborn, J.S.; Rauschel, F.M. *Bioorg. Chem*., **1988**, *16*, 206.
[25] Ratcliffe, R.G.; Sachar-Hill, Y. *Biol. Rev*., **2005**, *80*, 27.
[26] Knight, W. B.; Sem, D.S.; Smith, K.; Miziorko, H.M.; Rendina, A. R.; Cleland, W.W. *Biochemistry*, **1991**, *30*, 4970.
[27] Wińska, P.; Gołos, B.; Cieśla, J.; Zieliński, Z.; Frączyk, T.; Wałajtys-Rode, E.; Rode, W. *Parasitology*, **2005**, *131*, 247.
[28] http://bse.pnl.gov/bse/portal (accessed January 3rd, 2009, 15:00).
[29] Sadlej, A.J. *Collec. Czech. Chem. Commun*., **1988**, *53*, 1995.
[30] Sadlej, A.J.; Urban, M. *J. Mol. Struct.: Theochem*, **1991**, *234*, 147.
[31] Sadlej, A.J. *Theor. Chim. Acta*, **1992**, *81*, 45.
[32] Sadlej, A.J. *Theor. Chim. Acta*, **1992**, *81*, 339.

Bioorganic Chemistry 38 (2010) 74–80

Contents lists available at ScienceDirect

Bioorganic Chemistry

journal homepage: www.elsevier.com/locate/bioorg

Thiophosphorylation of free amino acids and enzyme protein by thiophosphoramidate ions

Tomasz Ruman [a,*], Karolina Długopolska [a], Agata Jurkiewicz [b], Dagmara Rut [a], Tomasz Frączyk [c], Joanna Cieśla [c], Andrzej Leś [b,d], Zbigniew Szewczuk [e], Wojciech Rode [a,c]

[a] Rzeszów University of Technology, Faculty of Chemistry, Department of Biochemistry and Biotechnology, 6 Powstańców Warszawy Ave., 35-959 Rzeszów, Poland
[b] University of Warsaw, Faculty of Chemistry, Quantum Chemistry Laboratory, 1 Pasteur Street, 02-093 Warsaw, Poland
[c] Nencki Institute of Experimental Biology, 3 Pasteur Street, 02-093 Warsaw, Poland
[d] Pharmaceutical Research Institute, 8 Rydygier St., 01-793 Warsaw, Poland
[e] University of Wroclaw, Faculty of Chemistry, ul. F. Joliot-Curie 14, 50-383 Wrocław, Poland

ARTICLE INFO

Article history:
Received 28 October 2009
Available online 26 November 2009

Keywords:
NMR
Thiophosphate
Tiophosphoramidate
Thiophosphorylation
Thymidylate synthase

ABSTRACT

In search of an activity-preserving protein thiophosphorylation method, with thymidylate synthase recombinant protein used as a substrate, potassium thiophosphoramidate and diammonium thiophosphoramidate salts in Tris- and ammonium carbonate based buffer solutions were employed, proving to serve as a non-destructive environment. Using potassium phosphoramidate or diammonium thiophosphoramidate, a series of phosphorylated and thiophosphorylated amino acid derivatives was prepared, helping, together with computational (using density functional theory, DFT) estimation of ^{31}P NMR chemical shifts, to assign thiophosphorylated protein NMR resonances and prove the presence of thiophosphorylated lysine, serine and histidine moieties. Methods useful for prediction of ^{31}P NMR chemical shifts of thiophosphorylated amino acid moieties, and thiophosphates in general, are also presented. The preliminary results obtained from trypsin digestion of enzyme shows peak at m/z 1825.805 which is in perfect agreement with the simulated isotopic pattern distributions for monothiophosphate of TVQQQVHLNQDEYK where thiophosphate moiety is attached to histidine (His^{26}) or lysine (Lys^{33}) side-chain.

1. Introduction

Physicochemical properties of phosphates allow them to play a dominating role in the living world [1]. Thiophosphate (or phosphorothioate) analogues, offering, as compared to the phosphate congeners, potential phosphorus-centered chirality and increased stability towards chemical and enzymatic hydrolysis [2–4], are often used in studies on biological activity of phosphorylated compounds [5–8]. The hydrolytic stability is of particular interest with studies on phosphorylation of protein basic amino acids, as due to their acid-labile character phosphoramidates within proteins easily escape detection by analytical methods [9,10]. Moreover, phosphohistidine, representing probably a major protein modification in eukaryotes [9], applied as a hapten has proved too unstable to generate antibodies [11]. In hope that thiophosphoramidates will be useful as more stable substitutes and probes for the biological function of the corresponding phosphoramidates, methods are sought allowing preparation of thiophosphoramidate-modified biomolecules under conditions mild enough to preserve biological function [12,13].

The first thiophosphorylation of peptide was described by Lasker et al. [12]. Surprisingly, despite the successful preparation of thiophosphohistidine by reaction with thiophosphoramidate, the authors were unable to thiophosphorylate histidine-rich peptide. Therefore they developed another method, employing $PSCl_3$ as a phosphorylation reagent.

The present study was done in search of an activity-preserving protein thiophosphorylation, with thymidylate synthase recombinant protein, found to undergo acid-labile phosphorylation [14], used as a substrate. Instead of phosphorus thiochloride that (i) shows high reactivity towards popular buffer ingredients, (ii) produces acidic conditions when used in water solutions and (iii) could act as a cross-linking agent giving products difficult to analyze [12], potassium thiophosphoramidate and diammonium thiophosphoramidate salts in Tris- and ammonium carbonate based buffer solutions were employed, proving to serve as a non-destructive environment. Using potassium phosphoramidate or diammonium thiophosphoramidate, a series of phosphorylated and thiophosphorylated amino acid derivatives was prepared, helping, together with computational (DFT) estimation of ^{31}P

* Corresponding author. Fax: +48 178543655.
E-mail address: tomruman@prz.edu.pl (T. Ruman).

0045-2068/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.bioorg.2009.11.002

NMR chemical shifts, to assign thiophosphorylated protein NMR resonances and prove the presence of thiophosphorylated lysine, serine and histidine moieties.

2. Results and discussion

2.1. NMR spectra of phosphorylated and thiophosphorylated amino acids

The ^{31}P NMR results for side-chain P- and TP-derivatives are presented in Table 1.

The most upfield shifted resonances (−5 to −10 ppm) among investigated phosphorylated amino acids belongs to phosphohistidines (P-His). Histidine reaction with phosphoramidates is significantly faster than in case of other amino acids but P–N bond that was created is vulnerable for acidic conditions as was found by Boyer group [15]. Histidine reaction with phosphoramidate produces two isomers 1-phosphohistidine and 3-phosphohistidine, the last one more thermodynamically stable. It should be noted that P-His isomers differ significantly in their hydrolysis rate, with 1-phosphohistidine having a 5 min hydrolytic half-life at pH 2.4 (46 °C), while 3-phosphohistidine has approx. five times longer half-life [13]. ^{31}P NMR spectrum of histidine and phosphoramidate post-reaction mixture contains two major resonances. The resonances at *upfield* region seem to belong to 1- and 3-phosphorylated histidine derivatives, with more *upfield* shifted resonance (−8 to −10 ppm) belonging to 1-phosphohistidine (δ_1-phosphohistidine) and the other one (−6 and −7 ppm) to 3-phosphohistidine (ε_2-phosphohistidine). ^{31}P NMR spectrum contains also two small singlet resonances belonging to 1,3-diphosphohistidine. The assignment of NMR resonances was based on previously published data [13].

The analogical reaction, carried out with thiophosphoramidate instead of phosphoramidate, led to histidine thiophosphates that are more hydrolytically stable than the corresponding phosphates, thus having significantly longer half-life [15]. The ^{31}P NMR spectrum of thiophosphorylated histidine contained resonances belonging to more thermodynamically stable 3-thiophosphohistidine isomer (3-TP-His; at 34.2 ppm), less stable 1-thiophosphohistidine (at 34.5 ppm) and two much smaller resonances belonging most likely to the 1,3-dithiophosphate species (at 33.7 and 34.4 ppm).

All phospho and thiophospho derivatives of cysteine, serine and lysine showed their resonances in the form of triplets, as an effect of the three-bond proximity of two magnetically equivalent hydrogen atoms. The coupling constants $^3J_{P-H}$ for those systems are in the 6–9 Hz range, in agreement with published data concerning similar compounds [16,17].

The most significant distinction between NMR properties of phosphate and thiophosphate compounds concerns large differences between chemical shifts. While the ^{31}P NMR resonances of P-His are in the −7 to −10 ppm range, those of TP-His are in the 33–35 ppm region. The magnitude of those differences may reflect a distinctly weaker electronic withdrawing effect of sulfur, as compared with oxygen which causes strong electronic deshielding of phosphorus. This *downfield* shifting effect is observable for all phosphorylated and thiophosphorylated amino acid pairs, and is especially visible for thiophosphocysteine, its two sulfur atoms connected to phosphorus and ^{31}P resonance shifted *downfield* to 54.2 ppm.

It should be noted that the averaged differences between ^{31}P NMR resonances of different pairs of P- and TP-amino acids are comparable, ranging from 43.2 ppm difference with 1-P-/1-TP-His, 43.4 ppm with P-/TP-Lys and 44.1 ppm with P-/TP-Ser to 41.0 ppm with 3-P-/3-TP-His and 38.3 ppm difference with P-/TP-Cys. Thus knowing the mean difference value, roughly 42 ppm, may be useful in prediction of ^{31}P NMR shifts of various thiophosphates when chemical shifts of the corresponding phosphates are known.

2.2. DFT calculations of ^{31}P shieldings of TP-amino acids

With an accurate and reliable estimation of chemical shifts for ^{31}P NMR method being still a long-term challenge, little is known in this respect about thiophosphates and their derivatives. While estimation of phosphorus chemical shift $\delta_{P,}$ based on comparison of electronic shielding of particular atoms, might provide very useful information regarding NMR properties of thiophosphates, it has so far poorly described.

Phosphorus electronic shielding calculations of side-chain thiophosphate derivatives of selected amino acids are presented in Table 2. The TP-amino acid models used in these calculations were divided into three groups based on molecular charge of model compound. All of the shielding values of TP-amino acids were converted into chemical shift values using analogically obtained

Table 1
^{31}P NMR chemical shifts of phosphorylated and thiophosphorylated amino acids.

P- or TP-amino acid	^{31}P NMR chemical shift [ppm] and/multiplicity	Averaged P-/TP-amino acid shift difference [ppm]
P-His[a]	−5 to −10/s	–
P-Arg	5–6/s	–
P-Lys	0–0.2/tr	–
P-Cys	15.9/tr	–
P-Ser	−0.3/tr	–
P-Tyr	−3.8–0.0/s	–
P-Thr	−1.4/m	–
TP-His[a]	32.5–34.5/s	43.2 (1-P-His)[e], 41.0 (3-P-His)[e]
TP-Arg	–	–
TP-Lys	43.4/tr	43.4[f]
TP-Cys	54.2/tr	38.3
TP-Ser	43.8/[b]	44.1
$PSCl_3$[c]	54.6/s, 49.7[d]; 28.6[d]; 15.7[d]; −0.1[d]/s	–

[a] Histidine phosphorylation product contain 3-P-His, 1-P-His and diphosphate (with phosphate groups both at 1- and 3-positions) with relative integrals being 13.9:1:0.5 respectively. Also the thiophosphorylation product contain 3-TP-His, 1-TP-His and analogical dithiophosphate with relative integrals being 8.5:1:0.14.
[b] Two doublets.
[c] Common trace contaminant of thiophosphoramidates.
[d] Phosphorus thiochloride hydrolysis products.
[e] Calculated using $\delta_{1\text{-P-His}} = -8.7$ ppm and $\delta_{3\text{-P-His}} = -6.8$ ppm.
[f] $\delta_{P\text{-Lys}}$ of 0.0 ppm was used for calculations (pH 7.5–7.8); higher values up to 0.2 ppm were observed at different pH values.

Table 2
Phosphorus isotropic shielding constants (σ_P, ppm) of side-chain thiophosphorylated amino acids estimated with the B3LYP/aug-cc-pVTZ GIAO calculations. δ_P stands for ^{31}P chemical shift in relation to H_3PO_4, $H_2PO_4^-$ or HPO_4^{2-}.

Neutral TP-amino acid	TP-Arg	TP-Cys	3-TP-His	TP-Lys	TP-Ser	TP-Thr
σ_P [ppm]	248.11	189.71	244.01	234.51	232.93	232.49
δ_P [ppm] rel. to H_3PO_4	62.62	121.02	66.72	76.22	77.80	78.24
δ_P [ppm] rel. to $H_2PO_4^-$	56.07	114.47	60.17	69.67	71.25	71.69
δ_P [ppm] rel. to HPO_4^{2-}	38.46	96.86	42.56	52.06	53.64	54.08
TP-amino acid monoanion						
σ_P [ppm]	255.65	223.81	257.59	255.88	254.84	245.13
δ_P [ppm] rel. to H_3PO_4	55.08	86.92	53.14	54.85	55.89	65.60
δ_P [ppm] rel. to $H_2PO_4^-$	48.53	80.37	46.59	48.30	49.34	59.04
δ_P [ppm] rel. to HPO_4^{2-}	30.92	62.76	28.98	30.69	31.73	41.44
TP-amino acid dianion						
σ_P [ppm]	276.55	216.67	255.33	266.40	256.36	244.65
δ_P [ppm] rel. to H_3PO_4	34.17	94.06	55.40	44.33	54.37	66.08
δ_P [ppm] rel. to $H_2PO_4^-$	27.63	87.51	48.85	37.78	47.82	59.53
δ_P [ppm] rel. to HPO_4^{2-}	10.02	69.90	31.24	20.17	30.21	41.92
TP-Cys 'thiol' form (P–SH) neutral		*TP-Cys 'thiol' form (P–SH) monoanion*				
σ_P [ppm]	261.38	σ_P [ppm]		275.16		
δ_P [ppm] rel. to H_3PO_4	49.35	δ_P [ppm] rel. to H_3PO_4		35.57		
δ_P [ppm] rel. to $H_2PO_4^-$	42.80	δ_P [ppm] rel. to $H_2PO_4^-$		29.02		
δ_P [ppm] rel. to HPO_4^{2-}	25.19	δ_P [ppm] rel. to HPO_4^2		11.41		

values of reference compounds – neutral H_3PO_4, monoanionic $H_2PO_4^-$ and HPO_4^{2-} dianion using the following equation:

$$\delta_P = \sigma_{ref} - \sigma_x \quad (\text{Ref. } [15])$$

where δ_P is phosphorus NMR chemical shift, σ_{ref} – isotropic phosphorus shielding of reference compound and σ_x – isotropic phosphorus shielding of analyzed compound.

It should be noted that precise estimation of phosphorus NMR chemical shift is still an unsolved and very difficult problem because of (i) unreliable estimation of concentration of neutral, mono- and dianionic forms in solution, (ii) complexity of interactions between analyzed compounds and buffer ingredients and also (iii) calculations of shieldings of those mentioned forms being still inaccurate. In our calculations the best match of obtained shift values with experimental ones was achieved with neutral TP-amino acid models (Table 2, Fig. 1). The experimental values of ^{31}P NMR chemical shifts for 3-TP-His (32.5 ppm), TP-Lys (43.4 ppm) and TP-Ser (43.8 ppm) could be correlated with neutral TP-amino acid models (related to H_3PO_4) by subtracting an additional factor of 33.7 ppm value. This subtraction yields the following values of correlated chemical shifts: 33.0 ppm (TP-His, −0.4 ppm difference), 42.5 ppm (TP-Lys, 0.88 ppm difference) and 44.1 ppm (TP-Ser, −0.30 ppm difference). These values are in a good agreement with experimental ones that showed ^{31}P NMR shifts of phosphates and thiophosphates to have their resonances within the range of almost 150 ppm. The computational prediction of ^{31}P chemical shift for TP-Arg and TP-Thr was also made (Table 2), suggesting chemical shifts of 28.9 ppm for TP-Arg and 44.5 ppm for TP-Thr.

Of interest is that the chemical shift values of TP-Cys phosphorus (see Table 2), as compared with those for TP-His, TP-Ser and TP-Lys, are much higher, although calculated in an identical manner. Furthermore, this value (87.3 ppm) differs considerably from experimental one (54.2 ppm). The explanation of this phenomenon was described previously by our research group [18]. The calculation results strongly suggest that TP-cysteine is existing in equilibrium of 'thione' tautomeric form with P=S formal double bond along with 'thiol' form with P=O formal double bond and ionized or protonated sulfur P–S^-/P–SH. Those 'thiol' tautomeric forms are having their ^{31}P NMR resonances *upfield* shifted by a factor of 20–30 ppm [18]. In order to estimate the chemical shift of 'thiol' form, we have conducted another set of calculations and those results are presented in Table 2 (bottom part). The calculated phosphorus chemical shift of neutral form of 'thiol' TP-Cys (in relation to H_3PO_4, with 33.7 ppm correlation factor) is approx. 16.7 ppm

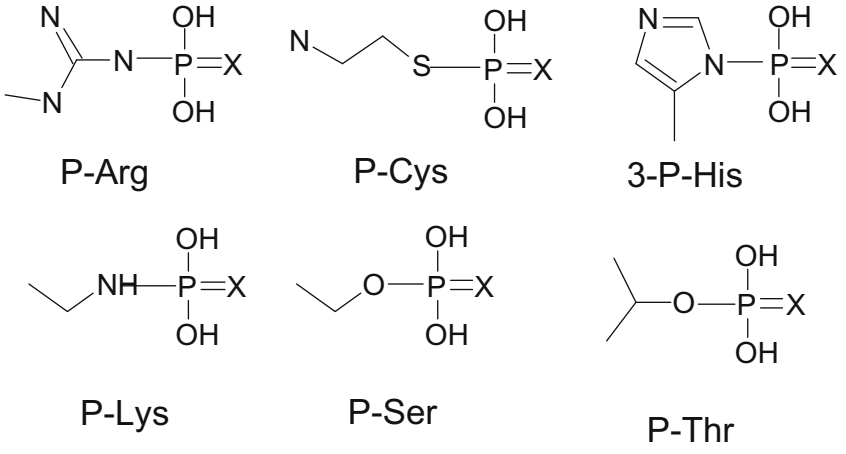

Fig. 1. Structures of the models of phosphorylated and thiophosphorylated amino acids. The symbol *X* denotes O or S atom.

being very similar to previously described thiophosphate 'thiol' forms [18]. The almost equimolar equilibrium of 'thione' (87.3 ppm) and 'thiol' (16.7 ppm) forms should give observed value of 54.2 ppm.

2.3. Thiophosphorylation of Caenorhabditis elegans thymidylate synthase

Fig. 2 presents thiophosphate 'fingerprint' region of ^{31}P NMR spectrum. Spectrum B at Fig. 2 presents results of the thiophosphorylation of Tris–HCl buffer, identical to one used for modification of thymidylate synthase. The thiophosphorylated Tris (fragment B at Fig. 2) presents two multiplets at approx. 42.3 and 43.9 ppm that are also observable in thiophosphorylated protein NMR spectrum (spectrum C at Fig. 2). The resonances of thiophosphorylated Tris at 42.3, 43.9 ppm (J_{P-H} = 7.6 Hz and 6.6 Hz respectively) are triplets as a result of heteronuclear coupling of thiophosphate phosphorus with two magnetically equivalent hydrogen atoms from neighboring methylene group [16]. Spectrum A in Fig. 2 presents the above described fingerprint region of thiophosphorylated protein with H-decoupling applied, showing clearly that thiophosphorylated Tris resonances are singlets, thus proving the three-bond P–H couplings.

The thiophosphorylated thymidylate synthase NMR spectrum is presented in Fig. 2, part C. Besides the thiophosphorylated Tris resonances, two triplets are present at approx. 43.2 and 43.3 ppm (coupling constants J_{P-H} = 6.5 and 5.6 Hz respectively) and two doublets at 43.8 ppm (J_{P-H} = 5.7 and 8.5 Hz). Coupling constants mentioned above are in the range typical for three-bond P–H coupling [19]. All of the resonances from thiophosphorylated protein are singlets with H-decoupling applied (Fig. 2, spectrum A), proving the three-bond proximity of thiophosphate phosphorus to the methylene hydrogen atoms. The resonance at approx. 43.8 ppm apparently belongs to the thiophosphorylated serine side-chain hydroxyl. The ^{31}P resonance of newly formed TP-Ser is in the form of two doublets, resulting from the heteronuclear coupling with phosphorus of two magnetically nonequivalent (shift difference 0.04 ppm) hydrogen atoms from TP-serine moiety methylene group. Those two magnetically nonequivalent hydrogen atoms have very similar coupling constants $^{3}J_{P-H}$ = 5.65 and 5.67 Hz. TP-Ser thesis is strongly supported by an almost identical chemical shift of TP-serine obtained by direct thiophosphorylation of serine (see Table 1) and ^{31}P–^{1}H gradient enhanced HMBC spectra pointing to 3.6 ppm as ^{1}H chemical shift of phosphorus-coupled hydrogen atoms.

Two triplet resonances at approx. 43.3 and 43.4 ppm correspond apparently to TP-Lys as thiophosphorylation of free amino acid lysine gave a very similar result (Table 1). The ^{31}P NMR spectrum of TP-Lys that was measured at identical pH as thiophosphorylated thymidylate synthase sample, contains a triplet resonance at 43.4 ppm. Also, the ^{31}P–^{1}H HMBC spectra confirm that phosphorus-coupling hydrogen atoms have ^{1}H NMR shifts of approx. 3.1 ppm, corresponding to ε-CH_2 group of lysine moiety. Another interesting observation that stems from the 2D spectrum is that the phosphorus atoms of both triplet resonances in TP-lysine region (43.3–43.4 ppm) are coupled to exactly the same ^{1}H chemical shift (3.1 ppm), suggesting those two triplets to reflect two forms of protein TP-lysine moiety resulting probably from a slow, at NMR-scale, protonation/deuteration of lysine ε-nitrogen atom caused by bulky thiophosphate substituent.

The results of our studies on free amino acid phosphorylation and thiophosphorylation pointed to histidine being the fastest reacting amino acid in the phosphorylation and thiophosphorylation reactions (not shown). Histidine thiophosphorylation reactions were performed with the free amino acid, as well as *C. elegans* recombinant thymidylate synthase protein (see Figs. 3 and 4, Table 1).

The free amino acid histidine thiophosphorylation reaction was carried out in ammonium carbonate buffer (pH 7.4). Series of resonances were found at 32.5–34.5 ppm region that is very near to that from an older work of Pirrung group showing approx 35 ppm chemical shift of TP-His [13]. However, that spectrum was obtained in deuterium oxide and no exact pH or pD value were presented; it should be noted that ^{31}P NMR chemical shifts of ions significantly depend on pH (or pD). The ^{31}P NMR spectrum of thiophosphorylated thymidylate synthase also presents two singlet resonances at 32.5 and 32.7 ppm laying in the TP-His region (Fig. 3). The more *downfield* shifted resonance has relative integral of 1.0 with another *upfield* shifted one of 1.7. The assignment of those two resonances was made by analysis of free histidine thiophosphorylation NMR results and limited literature data [13,15]. It seems that more *upfield* shifted resonance belongs to hydrolytically more stable 3-thiophosphohistidine (3-TP-His) and another one *downfield* shifted resonance from less hydrolytically stable 1-thiophosphohistidine (1-TP-His).

The exceptionally interesting part of the spectrum of thiophosphorylated thymidylate synthase is presented in Fig. 4. There are two singlet resonances at −6.2 and −6.3 ppm laying in region characteristic for phosphohistidines (P-His). Our free amino acid hisitidine phosphorylation results confirm that those two singlet resonances almost certainly belong to the 1-P-His and 3-P-His being most likely a hydrolysis product of 1-TP-His and 3-TP-His respectively. Relative integrals measured for those two resonances

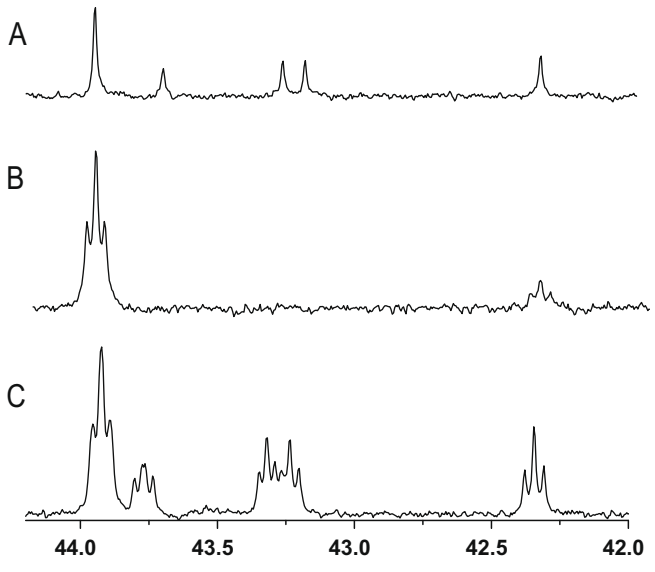

Fig. 2. Thiophosphate fingerprint region of ^{31}P NMR spectrum of thymidylate synthase thiophosphorylation product with ^{31}P–^{1}H decoupling (A) and without H-decoupling (C). Spectrum B shows thiophosphorylation product of Tris–HCl buffer.

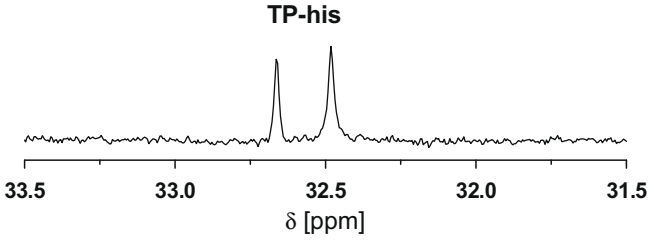

Fig. 3. Thiophosphohistidine resonances of ^{31}P NMR spectrum of thiophosphorylated thymidylate synthase.

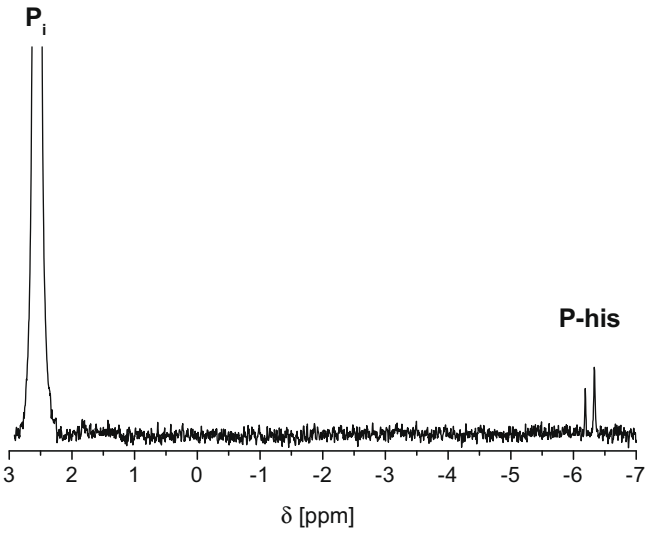

Fig. 4. Phosphohistidine resonances of ^{31}P NMR spectrum of thiophosphorylated thymidylate synthase. P_i = inorganic phosphate.

are 1:2.4 for 1-P-His and 3-P-His respectively. Thiophosphate may hydrolyze to phosphate in aqueous solutions, as evidenced by the time-dependent growth of the inorganic phosphate resonance at approx 2.2–2.4 ppm (see Fig. 4).

The enzyme thiophosphorylation was carried out also in ammonium carbonate buffer, pH 7.4, with the use of diammonium thiophosphoramidate. The results were similar to those obtained with Tris buffer but for (i) much lower total amount of thiophosphohistidines, (ii) higher 3-TP-His/1-TP-His molar ratio and (iii) much smaller corresponding resonances from TP-Ser and TP-Lys.

The mass spectrometry was also used to propose the location of the thiophosphate moiety in the thymidylate synthase molecule. The preliminary results obtained from trypsin digestion of enzyme shows peak at *m*/*z* 1825.805 The experimental isotopic patterns of the positively charged ion are in perfect agreement with the simulated isotopic pattern distributions for monothiophosphate of TVQQQVHLNQDEYK where thiophosphate moiety is attached to histidine (His^{26}) or lysine (Lys^{33}) side-chain (Fig. 1 in supplementary material).

Potential influence of the thiophosphorylation on thymidylate synthase activity was tested with recombinant human and *C. elegans* preparations of the enzyme. Each preparation underwent thiophosphorylation with diammonium TPA in Tris–HCl buffer, pH 7.5, at 4 °C under conditions described in Material and methods, with the enzyme activity determined in samples of the reaction mixture taken just after addition of diammonium TPA, and at several time points during 24 h period. Separate samples were incubated under the same conditions, but without diammonium TPA added, in order to test the enzyme stability. The results indicated the enzyme activity to be lowered, following 24 h reaction, to ~50% of the control value due to the thiophosphorylation process alone, pointing to a possibility that the modification influences the enzyme-catalyzed reaction parameters, as reported for phosphorylation [14].

3. Conclusions

The thiophosphorylation reactions of recombinant thymidylate synthase protein were performed using potassium thiophosphoramidate and diammonium thiophosphoramidate salts in Tris- and ammonium carbonate based buffer solutions at non-destructive environment. The NMR spectra of obtained modified proteins were assigned, based on NMR spectra of thiophosphorylated amino acid samples and the computational (DFT) estimation of ^{31}P NMR chemical shifts, indicating the presence of thiophosphorylated lysine, serine and histidine moieties. Several methods presented here should be useful for prediction of ^{31}P NMR chemical shifts of thiophosphorylated amino acid moieties and even thiophosphates in general.

4. Experimental

4.1. Materials and methods

Cellulose dialysis tubing, DE52-Cellulose, O-phosphotyrosine and O-phosphothreonine were purchased from Sigma–Aldrich. Diammonium- and potassium phosphoramidates (diammonium TPA and KTPA respectively) were prepared according to Pirrung et al. (2000).

4.2. NMR analysis

All NMR spectra were obtained with Bruker Avance spectrometer operating in the quadrature mode at 500.13 MHz for ^{1}H and 202.46 MHz for ^{31}P nuclei. The residual peaks of deuterated solvents were used as internal standards in ^{1}H NMR method. ^{31}P NMR spectra were recorded at 277 K both with and without proton decoupling. The internal standard used in ^{31}P NMR was inorganic phosphate (P_i) having its resonance at 2.15 ppm (at pH 7.8), 2.05 ppm (pH 7.5), 1.65 ppm (at pH 5.0) and 0.0 ppm (at pH 1.5). All samples were analyzed using gradient enhanced ^{1}H–^{31}P Heteronuclear Multiple Bond Correlation (HMBC) experiments. The HMBC experiments were optimized for long range couplings by using different $^{3}J_{P-H}$ values (1–20 Hz). The ^{1}H NMR spectra were obtained with the use of HDO suppression method. All buffer solutions used for NMR spectroscopy were based on deuterium oxide of 100%D purity (Armar Chemicals AG).

4.3. Mass spectrometry

The mass spectrum was obtained on a Bruker MicrOTOF-Q spectrometer (Bruker Daltonik, Bremen, Germany), equipped with Apollo II electrospray ionization source with ion funnel. The mass spectrometer was operated in the positive ion mode. The instrumental parameters were as follows: scan range *m*/*z* 200–2200, dry gas–nitrogen, temperature 180 °C, ion source voltage 4500 V, reflector voltage 1300 V, detector voltage 1920 V. The instrument was calibrated externally with the Tunemix™ mixture (Bruker Daltonik, Germany) in quadratic regression mode. The sample (0.1 M Tris–HCl buffer of pH 7.5 containing approx. 50 μg of protein and traces of thiophosphoramidate and thiophosphate salts) was diluted twice with methanol and infused at a flow rate of 3 μl/min. The mass spectrum was deconvoluted using the maximum entropy method.

The protein (0.2 mg) digestion was performed in the 0.2 M Tris–HCl buffer (pH 7.5) by addition of trypsin (10 μg) the reaction mixture was incubated at 24° C overnight. The solution was desalted on C18 Sep-pak Plus Cartridges (Waters, Ireland). Washed with water (10 ml) and eluted with acetonitrile (0.7 ml). The solution was concentrated *in vacuo* and introduced to mass spectrometer.

Analysis of the obtained spectra was performed with the Bruker Daltonics Compass DataAnalysis v.4 software The mass accuracy together with the isotopic pattern (using SmartFormula) enabled an unambiguous confirmation of the elemental composition of the obtained tryptic peptides.

4.4. Calculations

The theoretical calculations have been performed with the density functional B3LYP/aug-cc-pVTZ method. To save computational time, the structures of phosphorylated and thiophosphorylated amino acids were truncated by removing some groups of atoms remote from the phosphorylated regions (Fig. 1). The optimal geometries were obtained and confirmed with positive harmonic frequencies, then NMR shieldings for models of phosphorylated and thiophosphorylated amino acids in their neutral, monoanionic and dianionic forms were calculated. All the calculations were performed with the Gaussian G03 (rev. C.02) suite of programs [20].

4.5. Thiophosphorylation of amino acids (L-histidine, L-arginine, L-lysine, L-cysteine, L-serine, L-proline and L-tyrosine)

Ammonium carbonate buffer solution of amino acid (0.2 mmol of amino acid in 500 μl of buffer at pH 7.4) was reacted with diammonium thiophosphoramidate (1:3 amino acid to diammonium TPA molar ratios). Reaction mixture was shaken for 72 h at 277–278 K temperature. NMR sample was prepared by mixing 400 μl of post-reaction mixture with 300 μl of deuterium oxide. NMR results are presented in Table 1.

4.6. Phosphorylation of amino acids (L-histidine, L-arginine, L-lysine, L-cysteine, L-serine, L-proline, L-tyrosine and L-threonine)

Water solution of amino acid (0.2 mmol of amino acid in 500 μl of water) was reacted with potassium phosphoramidate (1–10 amino acid to KPA molar ratios). Reaction mixture was shaken for 48 h at 277–278 K temperature. NMR sample was prepared by mixing 100 μl of post-reaction mixture with 200 μl of deuterium oxide and 400 μl of Tris–HCl buffer (pH 7.5). NMR results are presented in Table 1.

4.7. Thymidylate synthase purified preparations and assay of activity

Human recombinant enzyme was expressed in thymidylate synthase-deficient *E. coli* TX61$^-$ strain (a kind gift from Dr. W.S. Dallas) as previously described [21], with the use of the plasmid construct pET17xb/hTS(LVAG) described by Pedersen-Lane et al., [22] and purified using the methods described for the rat enzyme [23]. *C. elegans* recombinant thymidylate synthase was expressed and purified as previously described [24]. The enzyme activity spectrophotometric assay [25] and protein content determination [26] were done by earlier described methods.

4.8. Thiophosphorylation of thymidylate synthase with potassium thiophosphoramidate (KTPA)

The reaction was started by introducing potassium thiophosphoramidate (42.0 mg) into 290 μl of the protein (1.064 mg) in 0.2 M Tris–HCl, pH 7.5. The reaction mixture was then shaken for 24 h at 277 K and analyzed using NMR spectrometer (300 μl of D_2O was added).

4.9. Thiophosphorylation of thymidylate synthase with diammonium thiophosphoramidate (diammonium TPA)

The reaction mixture was obtained by mixing 290 μl of the protein (1.065 mg) solution in 0.2 M Tris–HCl, pH 7.5, with diammonium thiophosphoramidate (41.0 mg). The reaction mixture was then shaken for 24 h at 277 K. The NMR sample contained obtained post-reaction mixture and 300 μl of deuterium oxide.

4.10. Thiophosphorylation of thymidylate synthase with diammonium TPA with buffer exchange

To the protein buffer solution (1 ml of 0.2 M Tris–HCl, pH 7.5 containing 2.1 mg of protein) diammonium TPA (23.0 mg) was added and shaken for 7 h at 277 K. Dialysis of the post-reaction mixture was then conducted by introducing obtained mixture into dialysis bag and dialyzed four times in ammonium carbonate buffer (4×300 ml of 0.1 M ammonium carbonate buffer, pH 7.4). NMR sample contained 0.5 ml of dialysis product and 0.3 ml of deuterium oxide.

4.11. Thiophosphorylation of thymidylate synthase with diammonium TPA in ammonium carbonate buffer

Diammonium TPA (13.4 mg) was introduced into protein solution (840 μg of *C. elegans* thymidylate synthase in 0.1 M ammonium carbonate buffer pH 7.4).

Acknowledgments

Supported by the Ministry of Science and Higher Education of Poland, Grant Nos. N401 024036 and N204 088 31/2052. The ICM computer center, University of Warsaw, Poland, is acknowledged for the G18-6 computer grant.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bioorg.2009.11.002.

References

[1] F.H. Westheimer, Science 235 (1987) 1173–1178.
[2] Z. Zhao, Biochem. Biophys. Res. Commun. 218 (1996) 480–484.
[3] L.A. Woźniak, A. Okruszek, Chem. Soc. Rev. 32 (2003) 158–169.
[4] J.J. Allen, S.E. Lazerwith, K.M. Shokat, J. Am. Chem. Soc. 127 (15) (2005) 5288–5289.
[5] F. Eckstein, Annu. Rev. Biochem. 54 (1985) 367–402.
[6] F. Eckstein, G. Gish, Trends Biochem. Sci. 14 (1989) 97–100.
[7] O. Heidenreich, W. Pieken, F. Eckstein, FASEB J. 7 (1993) 90–96.
[8] A. Mescalchin, A. Detzer, M. Wecke, M. Overhoff, W. Wünsche, G. Sczakiel, Expert Opin. Biol. Ther. 7 (2007) 1531–1538.
[9] H.R. Matthews, Pharmacol. Ther. 67 (1995) 323–350.
[10] S. Klumpp, J. Krieglstein, Biochim. Biophys. Acta 1754 (2005) 291–295.
[11] X.L. Zu, P.G. Besant, P.V. Attwood, Compr. Anal. Chem. 52 (2009) 315–352.
[12] M. Lasker, C.D. Bui, P.G. Besant, K. Sugawara, P. Thai, G. Medzihradszky, C.W. Turck, Protein Sci. 8 (10) (1999) 2177–2185.
[13] M.C. Pirrung, K.D. James, V.S. Rana, J. Org. Chem. 65 (2000) 8448–8453.
[14] T. Frączyk, T. Ruman, D. Rut, E. Dąbrowska-Maś, J. Cieśla, Z. Zieliński, K. Sieczka, J. Sikora, E. Wałajtys-Rode, D. Shugar, W. Rode, Anticancer Res. 28 (2008) 3462–3463.
[15] D.E. Hultquist, R.W. Moyer, P.D. Boyer, Biochemistry 5 (1966) 322–331.
[16] A.A. Isab, M.S. Hussain, M.N. Akhtar, M.I.M. Wazeer, A.R. Al-Arfaj, Polyhedron 18 (1999) 1401.
[17] J. Jankowska, A. Sobkowska, J. Cieślak, M. Sobkowski, A. Kraszewski, J. Skawiński, D. Shugar, J. Org. Chem. 63 (1998) 8150.
[18] T. Ruman, K. Długopolska, A. Jurkiewicz, D. Kramarz, T. Frączyk, A. Leś, W. Rode, The synthesis, reactivity and NMR investigation on^{15} N-thiophosphoramidates, Lett. Org. Chem. 6 (8) (2009).
[19] M.J. Gallagher, I.D. Jenkins, Topics in Stereochemistry, vol. 3, John Wiley & Sons, Inc., New York, 1968.
[20] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery Jr., T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople, Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford, CT, 2008.

[21] J. Cieśla, B. Gołos, E. Wałajtys-Rode, E. Jagielska, A. Płucienniczak, W. Rode, Acta Biochim. Pol. 49 (2002) 651–658.
[22] J. Pedersen-Lane, G.F. Maley, E. Chu, F. Maley, Protein Exp. Purif. 10 (1997) 256–262.
[23] J. Cieśla, K.X. Weiner, R.S. Weiner, J.T. Reston, G.F. Maley, F. Maley, Biochim. Biophys. Acta 1261 (1995) 233–242.
[24] P. Wińska, B. Gołos, J. Cieśla, Z. Zieliński, T. Frączyk, E. Wałajtys-Rode, W. Rode, Parasitology 131 (2005) 247–254.
[25] A.J. Wahba, M. Friedkin, J. Biol. Chem. 236 (1961) PC11–PC12.
[26] T. Spector, Anal. Biochem. 86 (1978) 142–146.

Organic & Biomolecular Chemistry

Dynamic Article Links

Cite this: *Org. Biomol. Chem.*, 2012, **10**, 323

www.rsc.org/obc

PAPER

Tyrosine nitration affects thymidylate synthase properties

Elżbieta Dąbrowska-Maś,†[a] Tomasz Frączyk,†[a] Tomasz Ruman,[b] Karolina Radziszewska,[c] Piotr Wilk,[a] Joanna Cieśla,[a] Zbigniew Zieliński,[a] Agata Jurkiewicz,[d] Barbara Gołos,[a] Patrycja Wińska,[a] Elżbieta Wałajtys-Rode,[e] Andrzej Leś,[d] Joanna Nizioł,[b] Adam Jarmuła,[a] Piotr Stefanowicz,[c] Zbigniew Szewczuk[c] and Wojciech Rode*[a,b]

Received 10th August 2011, Accepted 20th September 2011
DOI: 10.1039/c1ob06360j

Highly purified preparations of thymidylate synthase, isolated from calf thymus, and L1210 parental and FdUrd-resistant cells, were found to be nitrated, as indicated by a specific reaction with anti-nitro-tyrosine antibodies, suggesting this modification to appear endogenously in normal and tumor tissues. Each human, mouse and *Ceanorhabditis elegans* recombinant TS preparation, incubated *in vitro* in the presence of $NaHCO_3$, $NaNO_2$ and H_2O_2 at pH 7.5, underwent tyrosine nitration, leading to a V_{max}^{app} 2-fold lower following nitration of 1 (with human or *C. elegans* TS) or 2 (with mouse TS) tyrosine residues per monomer. Enzyme interactions with dUMP, meTHF or 5-fluoro-dUMP were not distinctly influenced. Nitration under the same conditions of model tripeptides of a general formula H_2N-Gly-X-Gly-COOH (X = Phe, Tyr, Trp, Lys, Arg, His, Ser, Thr, Cys, Gly), monitored by NMR spectroscopy, showed formation of nitro-species only for H-Gly-Tyr-Gly-OH and H-Gly-Phe-Gly-OH peptides, the chemical shifts for nitrated H-Gly-Tyr-Gly-OH peptide being in a very good agreement with the strongest peak found in ^{15}N-^{1}H HMBC spectrum of nitrated protein. MS analysis of nitrated human and *C. elegans* proteins revealed several thymidylate synthase-derived peptides containing nitro-tyrosine (at positions 33, 65, 135, 213, 230, 258 and 301 in the human enzyme) and oxidized cysteine (human protein Cys^{210}, with catalytically critical Cys^{195} remaining apparently unmodified) residues.

Introduction

Thymidylate synthase (TS; EC 2.1.1.45), a target in chemotherapy of a number of diseases, including cancer,[1] catalyzes the $N^{5,10}$-methylenetetrahydrofolate (meTHF)-assisted C(5)-methylation of dUMP,[2] required for DNA synthesis. It is, consequently, of interest to examine possible post-translational modifications of the enzyme in living cells.

Nitration of protein tyrosine residues is a post-translational modification, potentially affecting the function of a protein. It is associated with more than 50 diseases, including cancer, involving intensified NO biosynthesis.[3] The modification *in vivo* appears to be selective, with not many proteins becoming nitrated and only very few residues being modified in each protein. Moreover, even with good nitration targets, the yield of protein nitro-tyrosine formation is low.[4] Nevertheless, the few known examples show that nitration of one or two tyrosine residues is enough to cause loss or gain of function (for physicochemical consequences of protein tyrosine nitration, *cf*. ref. 4), suggesting a need for studies directed at protein structure-function analysis of specific proteins found to undergo nitration *in vivo*.[5]

As tetranitromethane nitration of sulfhydryl-blocked *Lactobacillus casei* thymidylate synthase protein caused enzyme inactivation,[6] it was of interest to test the possibility of TS tyrosine nitration in animal cells/tissue, and to determine to what extent enzyme properties might be affected by chemical nitration of TS tyrosine. Therefore kinetic and physicochemical (NMR, MS) studies were undertaken of *in vitro* nitrated preparations of human, mouse and *Caenorhabditis elegans* recombinant TSs. In order to enable interpretation of NMR resonances found in nitrated protein spectra, model tripeptides of a general formula H_2N-Gly-X-Gly-COOH (X = Phe, Tyr, Trp, Lys, Arg, His, Ser, Thr, Cys, Gly) were nitrated and analyzed using NMR spectroscopy. Besides, the theoretical calculations of ^{15}N NMR chemical shift for models of the nitrated tripeptides were performed and compared with those found experimentally.

[a]*Nencki Institute of Experimental Biology, Polish Academy of Sciences, Warszawa, Poland. E-mail: rode@nencki.gov.pl; Fax: (+48-22) 822 5342*
[b]*Rzeszów University of Technology, Faculty of Chemistry, Rzeszów, Poland*
[c]*Faculty of Chemistry, University of Wrocław, 14 F. Joliot-Curie Street, 50-383, Wrocław, Poland*
[d]*Faculty of Chemistry, University of Warsaw, 1 L. Pasteur Street, 02-093, Warszawa, Poland*
[e]*University of Information Technology and Management in Rzeszów, Chair of Cosmetology, 2 Sucharskiego Street, 35-225, Rzeszów, Poland*
† These authors contributed equally to this work.

Results and discussion

Reactivity of purified endogenous TS proteins to anti-tyrosine antibodies

Highly purified TS proteins, isolated from calf thymus, and L1210 parental and FdUrd-resistant cells, were found to be nitrated (Fig. 1), based on a specific reaction with anti-nitrotyrosine antibodies (Sigma-Aldrich, Anti-Nitro-tyrosine, Cat. No. N0409), suggesting the enzyme to undergo this modification endogenously in normal and tumor tissues. The reaction was specific to tyrosine, as the presence of nitro-tyrosine (10 mM) in the buffer containing anti-tyrosine antibodies (Fig. 1B), as well as the reduction of nitro-tyrosine to amino-tyrosine with $Na_2S_2O_4$ (Fig. 1C), prevented anti-tyrosine antibodies binding with the proteins. Of note is that the lack of the signal in panels B and C (Fig. 1) was not caused by an insufficient amount of protein, as demonstrated by the incubation of the same PVDF membrane with anti-TS antibodies (Fig. 1, E–F). It should be worthwhile to add that, according to the manufacturer, the polyclonal antibodies preparation used is fairly specific for protein nitro-tyrosine. While it recognizes nitrated proteins and 3-nitro-L-tyrosine, it does not cross-react with L-tyrosine, *p*-nitro-L-phenylalanine, 3-amino-L-tyrosine, 3-chloro-L-tyrosine and phospho-L-tyrosine BSA conjugates (http://www.sigmaaldrich.com/life-science/cell-biology/antibodies/learning-center/antibody-explorer/spotlights/anti-nitrotyrosine.html).

A comparison of nitration levels observed for the three endogenous enzyme preparations (Fig. 1, lanes 5–7) and that of the *in vitro* nitrated mouse recombinant enzyme determined to contain 0.8 and 1.6 mol/mol of TS monomer (Fig. 1, lanes 2 and 3, respectively), based on the ratio of signal intensities resulting from application of anti-nitro-tyrosine and anti-thymidylate synthase antibodies, indicates the modification level of the endogenous proteins to be much lower and concern presumably only a small fraction of each of those proteins.

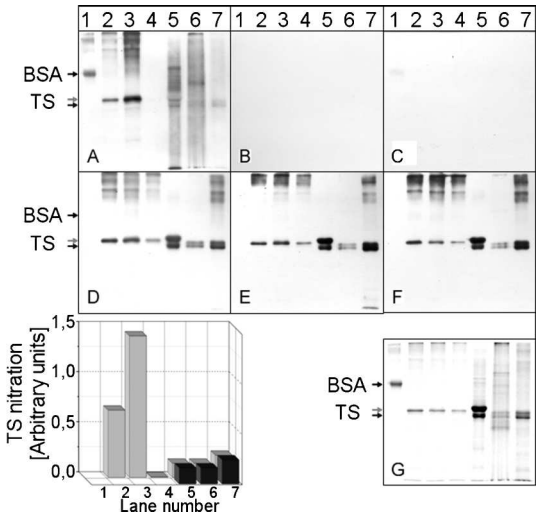

Fig. 1 Nitro-tyrosine detection by specific antibodies in chemically nitrated mammalian recombinant TS and endogenous enzyme preparations purified from tumour and normal tissues and separated by SDS-PAGE (without 2-mercaptoethanol). Proteins were stained with Sypro® Ruby Protein Gel Stain (G) or, following transfer to PVDF membrane, underwent first reaction with anti-nitroY antibodies (A–C), followed by removing of anti-nitroY antibodies and treatment with anti-TS antibodies (D–F). Negative controls included treatment with anti-nitroY antibodies either in the presence of 10 mM nitro-tyrosine (B) or following reduction of nitro-tyrosine to amino-tyrosine with 100 mM $Na_2S_2O_4$ at pH 9,0 (C). Nitrated BSA (positive control; lane 1), mouse recombinant TS nitrated with 8 mM (0.8 mol nitroY/mol TS subunit; lane 2), 12 mM (1.6 mol nitroY/mol TS subunit; lane 3) or 12 mM inactivated (negative control, 0 mol nitroY/mol TS; lane 4) peroxynitrite, and endogenous TS purified from calf thymus (lane 5), and L1210 parental (lane 6) and FdUrd-resistant (lane 7) cells. TS nitration presented in the bar chart was calculated as the ratio of signals deriving from bands of nitrated protein (A) and TS protein (D), with lighter and darker bars corresponding TS bands (marked with lighter and darker arrows in A and B) showing lower and higher mobility, respectively.

Recombinant TS *in vitro* nitration and its effect on enzyme properties

Each human, mouse and *Ceanorhabditis elegans* recombinant TS preparation, incubated *in vitro* at pH 7.5 in the presence of peroxynitrite ($ONOO^-$) producing mixture, containing $NaHCO_3$, $NaNO_2$ and H_2O_2 (1.00 : 1.05 : 1.00), underwent [H_2O_2]-dependent tyrosine nitration (Fig. 1, lanes 2–4, and Fig. 2; *Note*: H_2O_2 concentration reflects concentration of the peroxynitrite producing mixture, its constituents present always in constant proportion). The reaction dependence on H_2O_2 concentration varied in different TS proteins, being similar with the human (Fig. 2) and *C. elegans* proteins (linear progress observed in the range of 10–70 mM H_2O_2), but distinctly different with the mouse protein (linear progress observed in the range of only 5–20 mM H_2O_2, with lower nitration at higher concentrations). While this modification did not distinctly influence the K_m values reflecting enzyme interactions with dUMP and meTHF or the inhibition and inactivation rate constants (not shown) reflecting slow-binding of TS by 5-fluoro-dUMP (*cf.* ref. 7), it affected TS activity, leading to a V_{max}^{app} 2-fold lower following nitration of 1 (with human or *C. elegans* TS) or 2 (with mouse TS) tyrosine residues per monomer (not shown). It should be mentioned that initial experiments, involving TS nitration with synthesized authentic peroxynitric acid, showed the enzyme to undergo an instant inactivation (not shown), presumably due to the catalytic cysteine[2] modification. In accord, application of the peroxynitrite producing mixture, including CO_3^{2-}, known to inhibit sulfhydryl oxidation and enhance nitration of aromatics,[8] allowed to study nitration with the enzyme activity preserved.

NMR analyses of nitrated model compounds

In order to enable interpretation of resonances found in nitrated protein spectra, model compounds, including free amino acids (Phe, Tyr and Trp) and tripeptides of a general formula H_2N-Gly-X-Gly-COOH (X = Phe, Tyr, Trp, Lys, Arg, His, Ser, Thr, Cys, Gly), were nitrated and analyzed using NMR spectroscopy. While the three free amino acids did not allow quantitative analysis due to relatively low solubility of aromatic amino acids, such a study was possible with the use of the tripeptides, serving as simple models of proteins. The ^{15}N chemical shifts resulting from NMR studies, as well as DFT calculations performed for nitrated forms of truncated amino acid moieties (Fig. 3), are presented in Table 1. As the experimental NMR data clearly showed formation of nitro-species only for H-Gly-Tyr-Gly-OH and

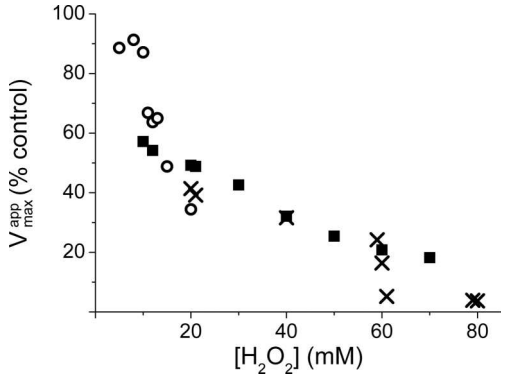

Fig. 2 Chemical nitration of human (squares), mouse (circles) and *C. elegans* (crosses) recombinant TS proteins: dependence of nitrated enzyme catalytic potency, reflected by the V_{max}^{app} value, on H_2O_2 concentration in the reaction mixture (Note: H_2O_2 concentration reflects concentration of the peroxynitrite producing mixture, its constituents present always in constant proportion). V_{max}^{app} value was determined at varying [dUMP] and constant [$N^{5,10}$-methylenetetrahydrofolate].

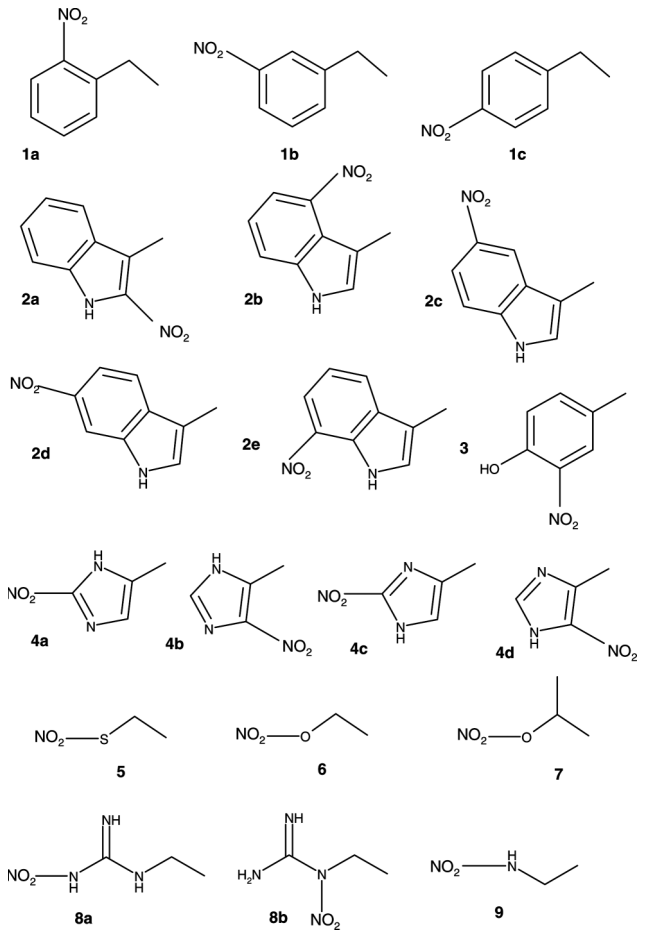

Fig. 3 Models of nitro-aminoacids used for DFT calculations: **1a** – *o*-NO_2Phe, **1b** – *m*-NO_2Phe, **1c** – *p*-NO_2Phe, **2a** – 2-NO_2Trp, **2b** – 4-NO_2Trp, **2c** – 5-NO_2Trp, **2d** – 6-NO_2Trp, **2e** – 7-NO_2Trp, **3** – 3-NO_2Tyr, **4a** – 2-NO_2-His N1-H, **4b** – 5-NO_2-His N1-H, **4c** – 2-NO_2-His N3-H, **4d** – 5-NO_2-His N3-H, **5** – *S*-NO_2Cys, **6** – *O*-NO_2Ser, **7** – *O*-NO_2Thr, **8a** – *N*-ω-NO_2Arg, **8b** – *N*-ε-NO_2Arg, **9** – *N*-NO_2Lys.

H-Gly-Phe-Gly-OH peptides, NMR-DFT comparison of chemical shifts was possible for these two systems. However, calculated ^{15}N chemical shifts, with chemical shift differences calculated as $\Delta\delta = \delta_{exp}{}^{15}N - \delta_{cal}{}^{15}N$ ($\delta_{exp}{}^{15}N$ – experimental chemical shift; $\delta_{cal}{}^{15}N$ – calculated chemical shift) for models of nitrated H-Gly-Tyr-Gly-OH and H-Gly-Phe-Gly-OH peptides (Table 1, column 5) taken into consideration, could be useful for prediction of ^{15}N chemical shifts of similar nitro-systems or even nitro-proteins. Of note is that recently Lehnik and Kirsch analyzed peroxynitrite nitration of L-tyrosine and related compounds with the use of ^{15}N CIDNP method.[9]

Nitration of the tyrosine-containing tripeptide resulted in the formation of 3-nitro-tyrosine moiety, along with only traces of other compounds (Fig. 4). Based on the ^1H NMR data, the molar ratio of tripeptide containing nitro-tyrosine to that containing tyrosine was 1 : 4.4. The confirmation of H-Gly-Tyr($^{15}NO_2$)-Gly-OH peptide formation was found in the ^{15}N-^1H HMBC spectrum (Fig. 5), clearly showing the heteronuclear coupling between ^{15}N-nitrate group (373.4 ppm) and the adjacent aromatic hydrogen atom at the ring C(2) (7.84 ppm). As indicated in Fig. 5, the ^{15}N resonance of the above mentioned nitro-moiety is a doublet, due to the 3-bond proximity of the hydrogen atom at the ring C(2).

Of note are interesting observations resulting from studies on the nitrated phenylalanine-containing tripeptide. The peaks observed in the ^{15}N-^1H HMBC spectrum (not shown) suggest the existence of two compounds, each containing nitro group attached to the aromatic ring. The NMR data, describing the cross peak, showing heteronuclear coupling between $^{15}NO_2$ nitrogen (371.2 ppm) and aromatic hydrogen atom (8.09 ppm), are in a very good agreement with those for 4-nitrotoluene that contains identical 4-nitrobenzyl moiety.[10] Thus the latter pair of chemical shifts points to the formation of H-Gly-Phe(NO_2)-Gly-OH peptide, with the $^{15}NO_2$ group generally connected to electronically and sterically favored *para* position.

The second cross peak (Fig. 5) found in ^{15}N-^1H HMBC spectrum (7.88 ppm at H-axis and 374.1 ppm) is in a surprisingly good agreement with our data found for H-Gly-Tyr(NO_2)-Gly-OH (*vide supra*). The NMR data strongly suggest the existence of nitro-tyrosine moiety that must have been formed by nitration of tyrosine moiety. Consequently, the tyrosine moiety must have resulted from phenylalanine moiety hydroxylation, a process described in the literature.[11,12] However, the yield of the overall process is very low, as the molar ratio calculated from ^1H NMR spectrum shows the amount of nitro-tyrosine-tripeptide to be almost six-fold lower than that of nitro-phenylalanine-tripeptide. The results obtained by ^{15}N NMR were validated by high resolution mass spectrometry (HRMS). The HRMS spectra revealed the presence of peaks corresponding to H-Gly-Phe($^{15}NO_2$)-Gly-OH (MNa^+ calculated 348.0932, found 348.0932), as well as H-Gly-Tyr($^{15}NO_2$)-Gly-OH (MNa^+ calculated 364.0882, found 364.0888). The intensities of the signals assigned to the nitration products, especially to H-Gly-Tyr(NO_2)-Gly-OH, were much lower than those of H-Gly-Tyr-Gly-OH and H-Gly-Phe($^{15}NO_2$)-Gly-OH.

The NMR data for nitrated H-Gly-Trp-Gly-OH peptide demonstrated only traces of nitrated forms as judged from ^{15}N (369.0 ppm) and ^1H (8.07 ppm) chemical shifts (not shown). The strongest cross peak in ^{15}N-^1H HMBC spectrum showed coupling between positions 400.7 ppm ^{15}N and 1.66 ppm ^1H of the spectrum but the chemical shift of the ^1H resonance (singlet resonance) can

Table 1 Experimental and calculated ^{15}N NMR chemical shifts of nitrated side chains of selected amino acid moieties

Residue	δ_{exp} ^{15}N (experimental; rel. to NH_3(lq)) [ppm][a]	δ_{exp} ^{1}H (experimental; H adjacent to N) [ppm]	δ_{cal} ^{15}N (calculated; rel. to NH_3(lq)) [ppm][b]	$\Delta\delta$[c] [ppm]	Nitro-product to substrate ratio[g]
o-NO_2Phe			382.8	—	—
m-NO_2Phe			389.7	—	—
p-NO_2Phe			389.7	–18.5	1 : 28.8
	371.2; 374.1[d]	8.09; 7.88[d]			1 : 156.3[d]
2-NO_2Trp	394.1[e]	2.78[e]	400.3	—	—
4-NO_2Trp	360.0[e]	1.70[e]	381.9	—	—
5-NO_2Trp	400.7[e]	1.66[e]	388.2	—	—
6-NO_2Trp			388.9	—	—
7-NO_2Trp			386.7	—	—
		7.84 (s, H-2)			
3-NO_2Tyr	373.4	7.37 (dd, H-6)	389.3	–15.9	1 : 4.4
		6.93 (d, H-5)			
2-NO_2-His N_1-H	—	—	408.0	—	—
5-NO_2-His N_1-H			396.6		
2-NO_2-His N_3-H			408.0		
5-NO_2-His N_3-H			403.6		
S-NO_2Cys		8.38 (s)[f]	370.0	—	—
O-NO_2Ser	—	8.39 (s)[f]	418.2	—	—
O-NO_2Thr	—	—	419.8	—	—
N-ω-NO_2Arg	—	—	429.1	—	
N-ε-NO_2Arg	—	—	416.9	—	
N-NO_2Lys	—	8.39 (s)[f]	407.7	—	

[a] Chemical shifts for nitrated tripeptides. [b] Calculated chemical shifts for truncated nitrated amino acids (Fig. 3). [c] $\Delta\delta = \delta_{exp}{}^{15}N - \delta_{cal}{}^{15}N$. [d] NO_2-Tyr (see Results and discussion). [e] See Results and discussion. [f] Side chain oxidation product. [g] Calculated from ^{1}H NMR spectrum.

not be associated with aromatic nitro-forms. It appears that the above mentioned singlet resonance reflects a tryptophan ring-cleavage product that contains an aliphatic region with ^{15}N atom.[12] Similar results, with only traces of nitro-form noticed (not shown), were obtained when post reaction mixture containing H-Gly-His-Gly-OH peptide was monitored. Moreover, the ^{15}N-^{1}H HMBC spectrum of the nitrated H-Gly-His-Gly-OH sample did not show any noticeable peaks corresponding to the ^{15}N and ^{1}H shifts of nitro-histidine.

The NMR spectra of nitrated serine and lysine peptides showed only traces of new forms. The ^{1}H NMR spectra of H-Gly-Cys-Gly-OH, H-Gly-Ser-Gly-OH and H-Gly-Lys-Gly-OH peptides contained small resonances at 8.4 ppm, suggesting oxidation of small percentage of peptides side chain heteroatoms to produce aldehydes. The ^{1}H spectra of nitrated threonine-containing peptide also showed traces of *O*-nitro-threonine moiety (CH–O–NO_2), with its ^{1}H resonance at 5.12 ppm and ^{15}N resonance at 535.5 ppm (not shown).

^{1}H NMR spectrum of nitrated H-Gly-Cys-Gly-OH showed almost all peptide used to be converted to various products. The literature suggests that the reaction of cysteine moiety with peroxynitrite may produce various amounts of disulfide (RSSR), sulfenic acid (RSOH), sulfinic acid (RSO_2H), sulfonic acid (RSO_3H), nitrosocysteine (RSNO), nitro-cysteine ($RSNO_2$) and various radicals.[12] Although the existence of RSNO and $RSNO_2$, containing ^{15}N-nitrogen atom originating from peroxynitrite, should be easy to confirm in 1D and 2D NMR spectra, it was not apparent. Consequently, the new products found in the ^{1}H NMR spectrum appear to be mainly disulfide, as well as sulfenic acid, sulfinic acid and sulfonic acid. The mass spectrometric analysis confirmed the presence of sulphonic acid in the sample.

NMR analyses of nitrated TS

The nitration of protein conducted under acidic conditions (pH 2–5) resulted in the precipitation of protein and consequently very low concentration of soluble ^{15}N-nitro-forms that could be analyzed by NMR method. The best results were achieved at pD of 7.2–7.4; when in NMR tube no presence of visible solid residues was apparent during the reaction and measurements. The chemical shifts for the strongest peak found in ^{15}N-^{1}H HMBC spectrum (7.67 ppm on the H-axis and 374.9 ppm on the ^{15}N axis) of nitrated protein (not shown) were in a very good agreement with our data for nitrated H-Gly-Tyr-Gly-OH peptide (*vide supra*). Due to relatively low concentrations of protein and very low yields of nitration process, it was impossible to confirm identity of other nitrated moieties.

MS analyses of nitrated TS

To localize the *in vitro* modifications of TS, nitrated human and *C. elegans* proteins underwent proteolytic digestion and the resulting peptides were analyzed by ESI. For each of the two proteins several TS-derived peptides were found to contain nitro-tyrosine (Table 2) and oxidized cysteine (Table 3) residues. It should be mentioned that although sequences of several peptides listed in Table 2 contained, besides tyrosine, also histidine residues, a possibility of their nitration appears negligible. The nitration of aromatic compounds is performed by the electrophilic aromatic substitution mechanism. Susceptibility of tyrosine side chain for electrophilic substitution is very high, in particular in its phenolate form at pH > 7. In contrast to this, the electron density of the imidazole ring of histidine is lower. Consequently, the electrophilic substitution of histidine residue is much slower and requires

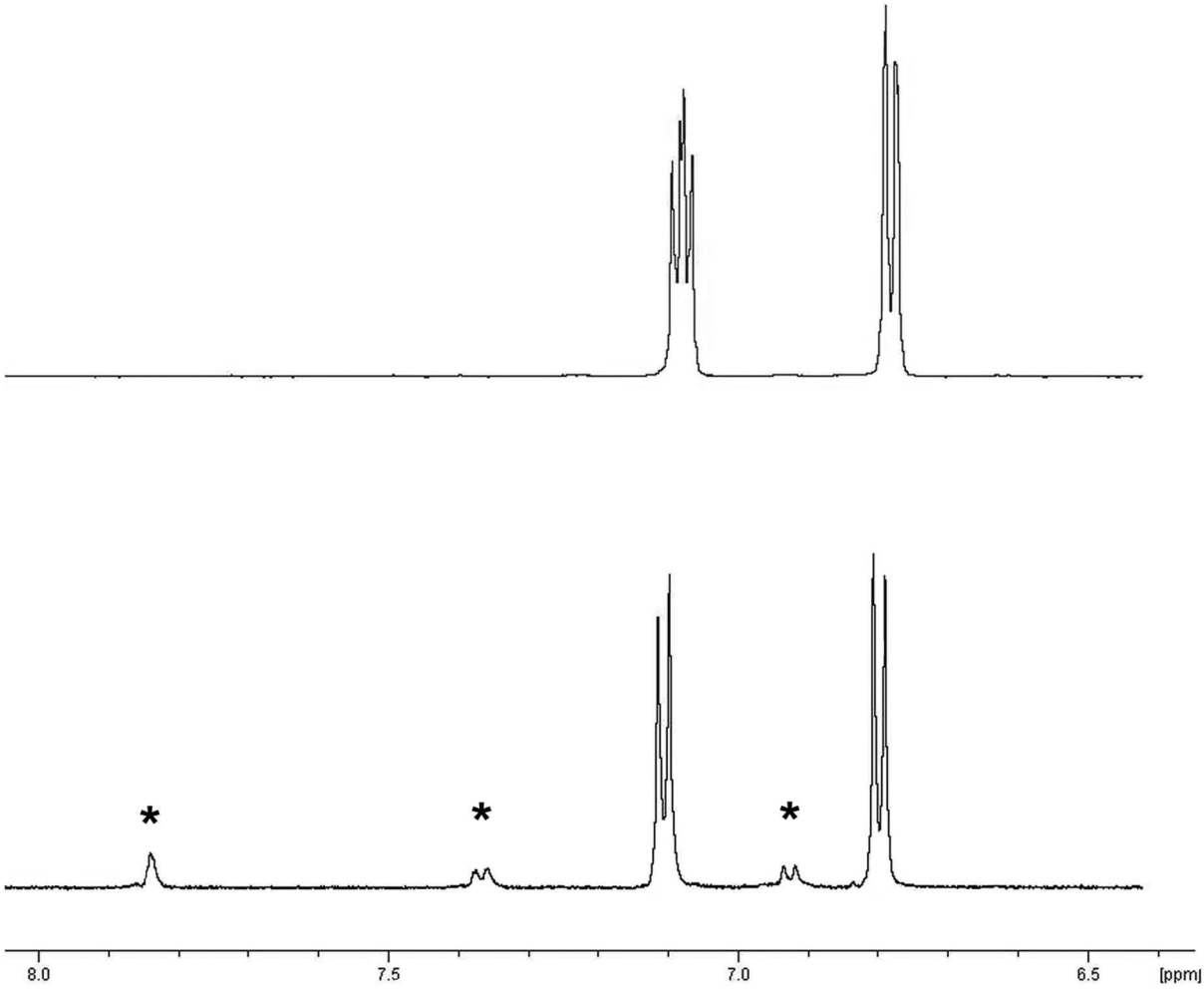

Fig. 4 Aromatic regions of H-Gly-Tyr-Gly-OH peptide (top) and post-reaction mixture of peptide ^{15}N-nitration (bottom). Asterisks indicate aromatic hydrogens of ^{15}N-nitro-tyrosine moiety.

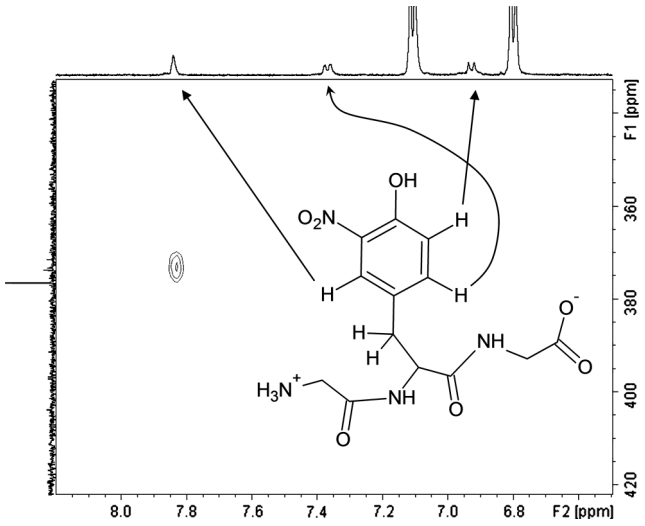

Fig. 5 The fragment of ^{15}N-^{1}H HMBC spectrum of the nitrated H-Gly–Tyr-Gly-OH peptide showing ^{15}N-H(Tyr) coupling. The largest peak on ^{15}N axis belongs to NO^{3-} anion.

higher pH than that of tyrosine. Therefore it appears hardly possible that in the presence of very reactive phenolate side chain the imidazole ring of histidine would be modified. In accord, ESI-MS analysis revealed only one nitration site in the 255–261 fragment of human thymidylate synthase, containing one tyrosine and two histidine residues (Fig. 6). Moreover, low susceptibility to nitration of histidine *vs.* tyrosine was shown also by the results of NMR analyses of nitrated model compounds (*vide supra*).

While the non-nitrated counterparts of several nitrated peptides could be identified, in most cases their proportion was probably too low to allow unambiguous recognition.

Following nitration at pH 7.5 with 20 mM H_2O_2, tyrosine modification to nitro-tyrosine was found at positions 33, 65, 135, 213, 230, 258 and 301 in human TS (Fig. 6 and 7), and 34, 66, 137, 148 and 232 (homologous to human 33, 65, 135, 146 and 230, respectively) in the *C. elegans* protein. Interestingly, when 50 mM H_2O_2 was applied with *C. elegans* TS, a different nitration profile, including positions 66, 137, 148, 204, 215 and 260 (homologous to human 65, 135, 146, 202, 213 and 258, respectively), was apparent. Of note is a tendency for homologous

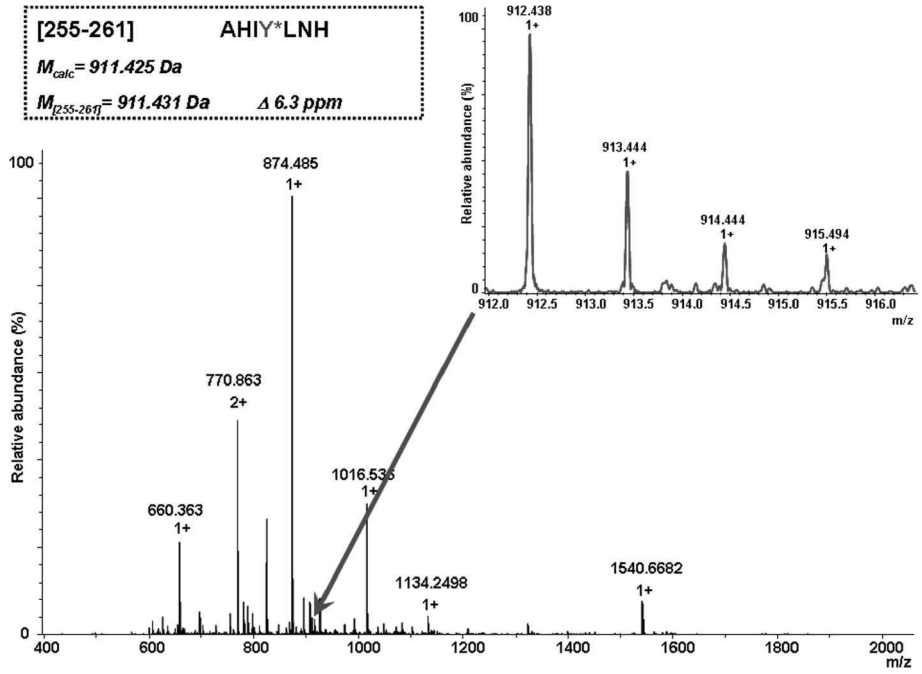

Fig. 6 The ESI-MS spectrum of the chymotryptic products of human TS. The expanded part of spectrum corresponds to the fragment 255–261 containing the nitrated tyrosine at position 258. The monoisotopic molecular mass of the presented fragment was compared with one simulated basing on the molecular formula.

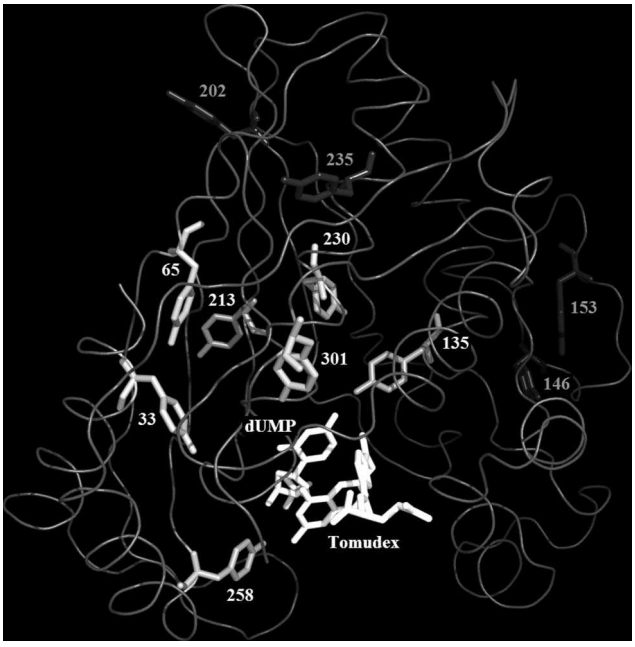

Fig. 7 Ribbon representation of monomer A from the crystal structure of human thymidylate synthase with bound dUMP and Tomudex (PDB code: 1I00). Tyrosine residues are shown as sticks, labelled with sequence numbers and marked either lighter or darker depending on whether a residue has been found to be nitrated or not, respectively. dUMP and Tomudex are shown as sticks and marked light.

tyrosine residues in both TS sequences to undergo nitration (Table 2). Considering localization of nitrated tyrosine residues in human TS, those at positions 135, 230 and 258, being closest to the active center (Fig. 7), were of particular interest, in view of the influence of nitration on the catalytic potency (Fig. 2). It should be mentioned that at the conditions used to nitrate the human enzyme (pH 7.5, 20 mM H_2O_2) the only cysteine residue to undergo modification was Cys^{210}, with catalytically critical Cys^{195} remaining apparently unmodified. In *C. elegans*, the homologous catalytic cysteine residue TS (Cys^{197}) underwent modification only at 50 mM H_2O_2, in accord with the dependence presented in Fig. 2, and cysteine residues at positions 67, 151 and 243, oxidized at 20 mM H_2O_2, correspond to human TS homologous amino acid residues different from cysteine (Ser^{66}, Met^{149} and Thr^{241}, respectively).

In order to asses a potential of nitration of different tyrosine residues to affect TS catalytic potency, a parallel molecular modeling study (using the molecular dynamics method, followed by post-processing of the resulting trajectories) has been performed. The simulations were based on the crystal structure of the ternary complex of human thymidylate synthase with dUMP and Tomudex (PDB accession code 1I00), with the Tomudex molecule replaced by the molecule of tetrahydrofolate (close analogue of methylenetetrahydrofolate), according to the previously described superimposition.[13] Initial results indicated nitration in human TS of either of the four residues, Tyr^{33}, Tyr^{135}, Tyr^{213} and Tyr^{258}, to differently influence the binding alignment between the substrate, dUMP, and cofactor, THF, in the enzyme active site. The impact ranges from (i) a strong misalignment that is likely to significantly reduce the catalytic activity of TS (nitration on Tyr^{135}) to (ii) moderate deviations from the native alignment (nitrations on Tyr^{33} and Tyr^{258}), and (iii) the lack of deviation from the native alignment (nitration on Tyr^{213}) that suggests preservation of TS native catalytic activity.[14] Considering the apparent strong influence of Tyr^{135} nitration on the substrates alignment, of interest is that mutation (Y94F) of the corresponding residue

Table 2 Mass spectrometric determination of tyrosine residues modification of $-NO_2$ type in human and *C. elegans* recombinant thymidylate synthase proteins nitrated *in vitro* at pH 7.5. Purified recombinant protein, following nitration, was analyzed by ESI-MS after proteolytic digestion. The peptides listed encompass those resulting from digestion by trypsin, as well as chymotrypsin

Sequence of the modified peptide	Modification site
Human TS protein nitrated at 20 mM H_2O_2	
A^{228}S**Y**ALL^{233}	Tyr^{230}
N^{205}SELSCQL**Y**QRSG^{217}	Tyr^{213}
F^{117}LDSLGFSTREEGDLGPV**Y**GFQW^{139}	Tyr^{135}
A^{255}HI**Y**LNH^{261}	Tyr^{258}
A^{22}EPRPPHGELQ**Y**LGQIQHILRCGV^{45}	Tyr^{33}
T^{251}LGDAHI**Y**^{258}	Tyr^{258}
S^{209}CQL**Y**QRSGDMGLGVPF^{225}	Tyr^{213}
K^{292}AEDFQIEG**Y**NPHPTIKM^{309}	Tyr^{301}
G^{60}MQAR**Y**SLRD^{69}	Tyr^{65}
C. elegans *TS protein nitrated at 20 mM $H_2O_2^-$*	
K^{33}**Y**LKQVE^{39}	Tyr^{34} (Tyr^{33})[a]
L^{133}GPV**Y**^{137}GFQW^{141}	Tyr^{137} (Tyr^{135})
F^{139}QWRHFGAK**Y**VDCHTD**Y**^{155}	Tyr^{148}; Tyr^{155}
G^{134}PV**Y**GFQWRHF^{144}	Tyr^{137} (Tyr^{135})
M^{221}GLGVPFNLAS**Y**GL^{234}	Tyr^{232} (Tyr^{230})
E^{274}P**Y**AFPK^{280}	Tyr^{276}
G^{61}MQSK**Y**CLRNG^{71}	Tyr^{66} (Tyr^{65})
C. elegans *TS protein nitrated at 50 mM H_2O_2*	
G^{61}MQSK**Y**CLRNG^{71}	Tyr^{66} (Tyr^{65})
Q^{40}ILREGTRRDDRTGTGTISIFGMQSK**Y**CLRNGTIPLLTTKRV^{81}	Tyr^{66} (Tyr^{65})
G^{190}QMVLPPCHTMCQF**Y**^{204}	Tyr^{204}
T^{126}SREEGDLGPV**Y**GFQW^{141}	Tyr^{137} (Tyr^{135})
A^{146}K**Y**VDCHTD^{154}	Tyr^{148}
Y^{204}VDNGELSCQL**Y**QRSGDMG^{222}	Tyr^{204}; Tyr^{215} (Tyr^{213})
G^{61}MQSK**Y**CLRNGTIPLLTTKRV^{81}	Tyr^{66} (Tyr^{65})
T^{56}ISIFGMQSK**Y**CL^{68}	Tyr^{66} (Tyr^{65})
T^{126}SREEGDLGPV**Y**GFQW^{141}	Tyr^{137} (Tyr^{135})
V^{251}HTLGDAHV**Y**^{260}	Tyr^{260} (Tyr^{258})

[a] Homologous human TS site, if also modified, is presented in parentheses.

in *E. coli* TS caused an apparent weakening of dUMP binding and associated enhancement of dUMP release, resulting in both substrates (dUMP and meTHF) interacting in a random binding sequence.[15]

In view of the recently presented concept of hydrogen bond bridges playing an important role in the reaction of protein tyrosine nitration,[16] it was of interest to extend our molecular modeling studies to analyze how the results of nitration of tyrosine residues in TS protein conformed to that concept. The results were unequivocal. While the distance between the nitrated tyrosine hydroxyl and the closest acidic/basic amino acid side chain heavy atom matches always satisfactorily the distance required for the nitrating species to form hydrogen bond bridge connecting the tyrosine and corresponding charged amino acid,[16] the corresponding distances measured to heteroatoms of the dissociable groups of the same amino acids are noticeably different from those suggested to be optimal,[16] and in accord, orientations between those dissociable groups and the nitrated tyrosine hydroxyls appear to be in most cases incorrect for forming geometrically reasonable hydrogen bond bridges. However, the latter statement should be taken with caution, as our molecular modeling was performed for the systems that had already undergone tyrosine nitration, hardly allowing evaluation of the intermediate conformational states occurring during nitration reactions.

Conclusions

The present study suggests that thymidylate synthase protein, expressed endogenously in normal and tumour (calf thymus and L1210 cells) tissues undergoes nitration *in vivo*. The modification may influence properties of the enzyme, as chemical reaction with peroxynitrite ($ONOO^-$) produced *in situ* at pH 7.5 of human, mouse and *C. elegans* recombinant TS proteins, resulting primarily in nitration of tyrosine residues, as confirmed by NMR and MS, distinctly lowers the catalytic potency reflected by the V_{max}^{app} value.

Experimental

Materials

Tripeptides of general formula H-Gly-X-Gly-OH where X = Phe, Tyr, Trp, Lys, Arg, His, Ser, Thr, Cys, Gly were purchased from Lipopharm (Poland).

Thymidylate synthase preparation

The endogenous enzyme proteins from parental and FdUrd-resistant mouse leukemia L1210 cells,[17] and calf thymus,[18] were purified as previously described. *Ceanorhabditis elegans*[19] and mouse[20] TS coding regions were cloned into pPIGDM4+stop vector and expressed as HisTag-free proteins in BL21(DE3)

Table 3 Mass spectrometric determination of cysteine residues modification of –O type in human and *C. elegans* recombinant thymidylate synthase proteins nitrated *in vitro* at pH 7.5. Purified recombinant protein, following nitration, was analyzed by ESI-MS after proteolytic digestion. The peptides listed encompass those resulting from digestion by trypsin, as well as chymotrypsin[a]

Sequence of the modified peptide	Modification site
Human TS protein nitrated at 20 mM H_2O_2	
N^{205}SELS**C**QLYQRSG^{217}	Cys^{210}
C. elegans *TS protein nitrated at 20 mM H_2O_2*	
G^{61}MQSKY**C**LRNG^{71}	Cys^{67}
F^{144}GAKYVD**C**HTDYSG^{157}	Cys^{151}
Q^{140}WRHFGAKYVD**C**H^{152}	Cys^{151}
R^{142}HFGAKYVD**C**HTDYSGQGVDQL^{163}	Cys^{151}
M^{238}IAKV**C**GLKPGTLVH^{252}	Cys^{243}
C. elegans *TS protein nitrated at 50 mM H_2O_2*	
G^{61}MQSKY**C**LRNG^{71}	Cys^{67}
Q^{40}ILREGTRRDDRTGTGTISIFGMQSKY**C**LRNGTIPLLTTKRV^{81}	Cys^{67}
G^{190}QMVLPPCHTM**C**QFY^{204}	Cys^{201}
A^{146}KYVDCHTD^{154}	Cys^{151}
K^{171}EQPDSRRIIMSAWNPSDLGQMVLPP**C**HTM**C**QFYVDNGE^{209}	Cys^{197}; Cys^{201}
G^{190}QMVLPP**C**HTM**C**QFYVDNGELS**C**QL^{214}	Cys^{197}; Cys^{201}; Cys^{212} (Cys^{210})

[a] Homologous human TS site, if also modified, is presented in parentheses.

or thymidylate synthase-deficient TX61$^-$ (a kind gift from Dr W. S. Dallas) *E. coli* strain, respectively. Human TS coding regions[21] were subcloned into pET28a vector and expressed as HisTag-containing proteins in *E. coli* BL21(DE3) strain. HisTag containing proteins were purified on NiNTA His-Bind resin (Novagen) according to manufacturer protocol, and HisTag-free proteins were purified as previously described.[22] Phosphatase inhibitors (50 mM NaF, 5 mM Na-pyrophosphate, 0.2 mM EGTA, 0.2 mM EDTA and 2 mM Na_3VO_4) were present in the purification buffers. Each purified TS preparation was separated from phosphorylated fraction according to Wolschin *et al.*,[23] using metal oxide/hydroxide affinity chromatography on $Al(OH)_3$ beads. The enzyme activity was measured and kinetic parameters of the enzyme-catalyzed reaction were determined as previously described.[24]

Thymidylate synthase tyrosine *in vitro* nitration

The reaction was performed at 4 °C in the presence of 20 μM dUMP (stabilization in the absence of 2-mercaptoethanol) in a reaction mixture containing 200 mM Na/K phosphate buffer pH 7.5, equimolar concentration of $NaHCO_3$ and H_2O_2 (5–70 mM), $NaNO_2$ at concentration by 5% exceeding the latter (5.25–73.5 mM) and the enzyme (5 μM dimer). To start the reaction, H_2O_2 was added, the sample mixed 30 s and next incubated 5 min. While nitro-tyrosine content was determined spectrophotometrically,[25] to the remaining reaction mixture 2-mercaptoethanol (20 mM) was added, followed by either protein precipitation with 10% (w/v) TCA or sample dilution (≥300-fold) with 50 mM Na/K phosphate buffer pH 7.5, containing 0.1% Triton X-100 and 10 mM 2-mercaptoethanol. The diluted preparation preserved TS activity for at least 2 h, allowing enzyme properties to be studied. To the control reaction mixture TS was added after mixing and incubating the remaining components, in order to inactivate the produced peroxynitrite.

Immunoblotting

Previously described method was used,[13] with anti-tyrosine antibody from Sigma–Aldrich (Cat. No. NO409) and anti-TS antibody.[19]

Peptide nitration

Peptides of the general formula H-Gly-X-Gly (where X = Phe, Thr, Trp, Lys, Cys, His, Ser, Arg, Gly) were nitrated in deuterium oxide system using sealed 5 mm NMR tubes. Each tripeptide (8 μmol in 0.1 ml of deuterium oxide) was mixed with solutions containing (i) 76 μmol H_2O_2 in 0.1 ml of D_2O, (ii) 80 μmol sodium ^{15}N-nitrite in 0.1 ml of D_2O and (iii) sodium bicarbonate (80 μmol in 0.1 ml D_2O), followed by addition of 3.3 μl of conc. sulfuric acid in 0.1 ml of D_2O and additional D_2O to the final volume of 0.6 ml (resulting pD of 1.9).

NMR analyses

All NMR spectra were obtained with Bruker Avance spectrometer operating in the quadrature mode at 500.13 MHz for ^{1}H and 50.69 MHz for ^{15}N nuclei. The residual peaks of deuterated solvents were used as internal standards in ^{1}H NMR method. ^{15}N NMR spectra were recorded at 277 K both with and without proton decoupling. All ^{15}N chemical shifts presented in this work are related to liquid ammonia (0.0 ppm). The internal standard used in ^{15}N NMR was sodium nitrite (609.6 ppm rel. to liquid NH_3) and sodium nitrate (376.5 ppm rel. to liquid NH_3). All samples were analyzed using the gradient-enhanced ^{1}H-^{15}N Heteronuclear Multiple Bond Correlation (HMBC) approach. The ^{1}H NMR spectra were obtained with the use of the HDO suppression method. All buffer solutions used for NMR spectroscopy were based on deuterium oxide of '100%D' purity (Armar Chemicals AG, Germany).

Calculations

The theoretical calculations have been performed with the density functional B3LYP/aug-cc-pVTZ method. To save computational time, the structures of amino acids were truncated by removing some groups of atoms remote from the nitrated regions (Fig. 3). Where it was possible, the calculations for full nitro-tripeptide (GCG-NO2) were performed for comparison with the truncated model. The optimal geometries were obtained and confirmed with positive harmonic frequencies, then NMR shielding values for models of nitroamino-derivatives of acids were calculated. All the calculations were performed with the Gaussian G03 (rev. C.02) suite of programs.

Mass spectrometry analysis

Enzymatic hydrolysis. A sample of modified protein (1 mg) was dissolved in water (100 µl). The obtained solution (10 µl) was diluted with 0.03 M NH_4HCO_3 (50 µl). After adding the 0.1% water solution of the proteolytic enzyme (5 µl; trypsin or chymotrypsin) the mixture was incubated at r.t. for 12 h. The products of the hydrolysis were absorbed on OMIX C4 100 µl pipette tips (Varian). The tip was washed 10 times with water (100 µl), than the peptides formed by proteolysis of modified protein were eluted with 60% water solution of acetonitrile (100 µl).

Mass spectrometry. The enzymatic digest was analyzed on high resolution ESI-FT mass spectrometer (Apex-Ultra Qe 7T; Bruker Daltonics; Germany) equipped with an electrospray ionization (ESI) source. The instrument was operated both in the positive and negative ion mode and calibrated with the Tunemix mixture (Bruker Daltonics). The mass accuracy was better than 5 ppm. The instrumental parameters were as follows: scan range, 300–2500 m/z; drying gas, nitrogen; temperature of drying gas, 200 °C; potential between spray needle and orifice, set at 4.5 kV; source accumulation time, 0.5 s; and ion accumulation time, 0.5 s. The analyzed solution was infused directly to the ion source at a flow rate 3 µl min^{-1}. Each spectrum is an average of more than 100 individual scans.

Data analysis. The analysis of spectra was performed using the SNAP algorithm (Data Analysis, Bruker). The generated mass list, including m/z ratio, monoisotopic mass and z, was further analyzed using Excel spreadsheet basing on the following assumptions: (i) The cleavage sites for trypsin: [K,R] and for chymotrypsin [Y,W,F,L,M,H,N,G,I,V,E,D]; (ii) accepted chemical modifications nitration of tyrosine and oxidation of the methionine and cysteine; (iii) Accepted error max. 10 ppm; (iv) Peptide length max. 20 amino acid residues; (v) Only protonated peptides were accepted, the metal adducts (Na, K, Ca) were neglected.

Acknowledgements

Supported by the Ministry of Science and Higher Education (grant numbers N401 0612 33 and N N401 0240 36). Stimulating discussions from COST Action CM1001 are acknowledged.

References

1 N. L. Lehman, *Expert Opin. Invest. Drugs*, 2002, **11**, 1775–1787.
2 C. W. Carreras and D. V. Santi, *Annu. Rev. Biochem.*, 1995, **64**, 721–762.
3 K. S. Aulak, M. Miyagi, L. Yan, K. A. West, D. Massillon, J. W. Crabb and D. J. Stuehr, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 12056–12061.
4 J. M Souza, G. Peluffo and R. Radi, *Free Radical Biol. Med.*, 2008, **45**, 357–366.
5 H. Ischiropoulos, *Biochem. Biophys. Res. Commun.*, 2003, **305**, 776–783.
6 D. Rosson, H. B. Otwell and R. B. Dunlap, *Biochem. Biophys. Res. Commun.*, 1980, **97**, 500–505.
7 K. Felczak, A. Miazga, J. Poznański, M. Bretner, T. Kulikowski, J. M. Dzik, B. Gołos, Z. Zieliński, J. Cieśla and W. Rode, *J. Med. Chem.*, 2000, **43**, 4647–4656.
8 B. Alvarez, G. Ferrer-Sueta, B. A. Freeman and R. Radi, *J. Biol. Chem.*, 1999, **274**, 842–848.
9 M. Lehnig and M. Kirsch, *Org. Biomol. Chem.*, 2006, **4**, 721–729.
10 Aldrich NMR Library, also available at www.aldrich.com (accessed 18.2.2011).
11 A. Van der Vliet, C. A. O'Neill, B. Halliwell, C. E. Cross and H. Kaur, *FEBS Lett.*, 1994, **339**, 89–92.
12 B. Alvarez and R. Radi, *Amino Acids*, 2003, **25**, 295–311.
13 A. Jarmuła, T. Frączyk, P. Cieplak and W. Rode, *Bioorg. Med. Chem.*, 2010, **18**, 3361–3370.
14 A. Jarmuła, unpublished work.
15 B. Hong, F. Maley and A. Kohen, *Biochemistry*, 2007, **46**, 14188–14197.
16 A. S. Bayden, V. A. Yakovlev, P. R. Graves, R. B. Mikkelsen and G. E. Kellogg, *Free Radical Biol. Med.*, 2011, **50**, 749–762.
17 J. Cieśla, T. Frączyk, Z. Zieliński, J. Sikora and W. Rode, *Acta Biochim. Pol.*, 2006, **53**, 189–198.
18 W. Rode, J. Cieśla, Z. Zieliński and B. Kędzierska, *Int. J. Biochem.*, 1986, **18**, 361–368.
19 P. Wińska, B. Gołos, J. Cieśla, Z. Zieliński, T. Frączyk, E. Wałajtys-Rode and W. Rode, *Parasitology*, 2005, **131**, 247–254.
20 J. Cieśla, B. Gołos, E. Wałajtys-Rode, E. Jagielska, A. Płucienniczak and W. Rode, *Acta Biochim. Pol.*, 2002, **49**, 651–658.
21 J. Pedersen-Lane, G. F. Maley, E. Chu and F. Maley, *Protein Expression Purif.*, 1997, **10**, 256–262.
22 J. Cieśla, K. X. Weiner, R. S. Weiner, J. T. Reston, G. F. Maley and F. Maley, *Biochim. Biophys. Acta, Gene Struct. Expression*, 1995, **1261**, 233–242.
23 F. Wolschin, S. Wienkoop and W. Weckwerth, *Proteomics*, 2005, **5**, 4389–4397.
24 W. Rode, T. Kulikowski, B. Kędzierska and D. Shugar, *Biochem. Pharmacol.*, 1987, **36**, 203–210.
25 J. P. Crow and H. Ischiropoulos, *Methods Enzymol.*, 1996, **269**, 185–194.